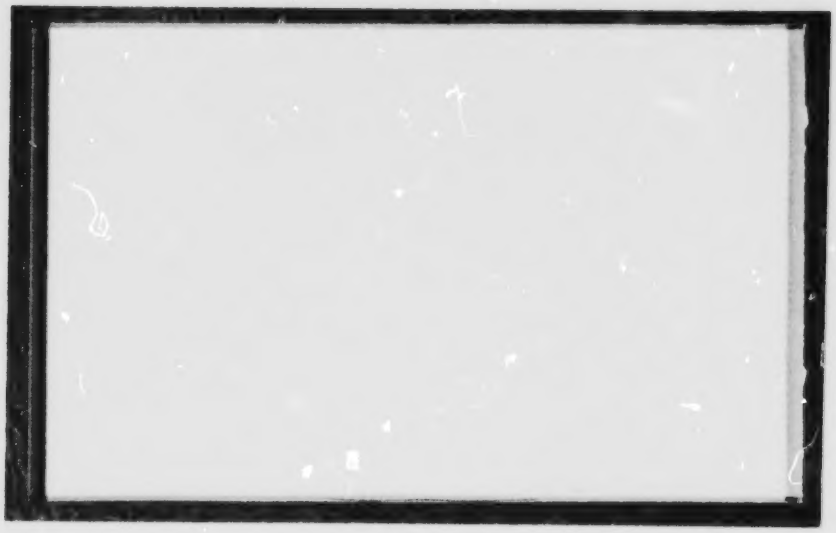
UNIVERSITÉ DE MONTRÉAL



CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

METHODES STATISTIQUES
EN OPTIMISATION COMBINATOIRE

par
Christian Lardinois

Ce travail a été subventionné par le Centre de recherche et de développement de Transports Canada, par le programme F.C.A.C. du Ministère de l'éducation du Québec ainsi que par le Prêt d'Honneur Inc.

Université de Montréal
Centre de recherche sur les transports - Publication # 116
Décembre 1978 - Révisé en avril 1979

Dépôt légal - 2e trimestre 1979

RESUME

L'approche développée dans ce rapport se situe au point de rencontre de la statistique et de l'optimisation combinatoire. Son objectif premier est l'estimation de l'optimum global des problèmes de grande taille; cette estimation est nécessaire pour juger de la qualité de toute solution heuristique, c'est-à-dire pour évaluer l'écart entre la valeur d'une telle solution et l'optimum global en question. Le rapport présente une synthèse des estimateurs proposés par différents chercheurs; l'accent est mis sur un estimateur qui provient de l'estimation de la distribution statistique des optima locaux associés à un algorithme de hill-climbing. Dans le cadre de cette approche, un nouvel intervalle de confiance pour l'optimum global est proposé. Il est ensuite proposé d'étendre cette procédure d'estimation au cas où les optima locaux sont relatifs à des heuristiques différentes. Un autre estimateur issu cette fois de l'estimation de la distribution des extrema d'échantillons d'optima locaux est également proposé.

Le rapport présente ensuite une règle itérative qui, d'une part, définit le nombre d'optima locaux qu'il est judicieux d'engendrer successivement, et qui d'autre part précise les conditions d'arrêt de l'exploration.

Enfin, l'utilisation de l'estimation de l'optimum global comme borne supérieure initiale dans un algorithme de branch and bound est proposée; ou bien on engendre ainsi la solution exacte, ou bien on obtient une solution heuristique meilleure que celles précédemment connues, ou bien encore aucune solution n'est produite; dans ces deux derniers cas, de nouveaux intervalles de confiance à 100% pour la solution exacte sont déduits.

TABLE DES MATIERES

INTRODUCTION. 1

1. POSITION DU PROBLEME 4
   1.1. La nécessité des techniques heuristiques. 4
   1.2. Notations . Domaine d'application. 6
   1.3. Les méthodes de hill-climbing 8

2. ESTIMATION DE L'OPTIMUM GLOBAL: ETAT DE L'ART 14
   2.1. Remarque préliminaire 14
   2.2. Inférence statistique classique et génération aléatoire 15
   2.3. Les premiers travaux de Dannenbring et ceux de Robson-Whitlock. 18
   2.4. Les premières recherches inspirées de Fisher-Tippett. 21
   2.5. Les recherches de Golden et Alt 26
   2.6. Critique de Golden et Alt 31

3. DISTRIBUTIONS PROPREMENT DITE ET INDUITE 35

4. PROLONGEMENTS. 44
   4.1. Intervalle de confiance. 44
   4.2. Utilisation d'heuristiques différentes. 49
   4.3. Passage à un niveau"supérieur". 50

5. RECAPITULATION ET EVALUATION DES ESTIMATIONS 52
   5.1. Estimations pour un problème donné. 52
   5.2. Comparaison et évaluation des estimateurs 56

6. REGLE D'ARRET. 59
   6.1. Une règle d'arrêt élémentaire 60
   6.2. La probabilité d'amélioration 60
   6.3. La valeur espérée du gain et la règle de McRoberts. 62
   6.4. Elaboration d'une règle générale intégrant les trois critères. . 63
      1) Calcul de la probabilité d'amélioration sur k observations additionnelles. 64
      2) Calcul de la valeur espérée du gain sur k observations additionnelles. 64
      3) Description de la procédure. 65

7. UN ALGORITHME HEURISTIQUE A PRIORI, PARFOIS EXACT A POSTERIORI . . . 68
   7.1. Description de la procédure 68
   7.2. Calcul d'intervalles de confiance 75

CONCLUSION. 78

ANNEXE 1: TEST DE KOLMOGOROV-SMIRNOV 81

ANNEXE 2: TEST DE CARACTERE ALEATOIRE. 84

BIBLIOGRAPHIE 86

INTRODUCTION

Pour résoudre les problèmes d'optimisation combinatoire de grande taille, on doit très souvent avoir recours à des procédures heuristiques, des procédures qui produisent en général de très bonnes solutions, mais qui ont cependant la particularité de ne fournir aucune garantie quant à la qualité de ces solutions. Ce papier a précisément pour premier objectif l'évaluation de cette qualité, c'est-à-dire l'évaluation de l'écart entre la valeur d'une solution heuristique et la valeur de la solution exacte. Dans ce but, le seul moyen est de réaliser une estimation de cette solution exacte[1]. C'est l'objet de la deuxième section que de présenter une synthèse des méthodes qui ont été imaginées pour estimer la solution exacte des problèmes combinatoires de grande taille; toutes ces méthodes procèdent de la statistique théorique; la plus prometteuse est fondée sur l'estimation de la distribution statistique des solutions heuristiques obtenues à l'aide d'un type particulier d'algorithme, à savoir un algorithme de "hill-climbing"; les solutions obtenues avec un tel algorithme sont des "optima locaux". La troisième section fait la distinction entre la distribution statistique caractérisant l'ensemble des optima locaux associés à l'algorithme de hill-climbing, et la loi de probabilité décrivant la distribution des optima susceptibles d'être engendrés, laquelle distribution est induite par le hill-climbing et la procédure de génération aléatoire des solutions initiales nécessaires à ce hill-climbing. Trois extensions sont ensuite proposées; la première définit un nouvel intervalle de confiance pour la solution exacte; la seconde suggère l'utilisation des optima locaux engendrés par des

---

1. Plus exactement: une estimation de la valeur de cette solution... (voir paragraphe 1.2.)

algorithmes différents; la troisième fait porter l'attention sur les extrema d'échantillons d'optima locaux. Ces prolongements sont présentés dans la quatrième section. La cinquième section récapitule ces différentes estimations et indique comment les évaluer.

Le problème de la proximité du meilleur optimum local atteint (à un moment donné) avec un algorithme de hill-climbing vis-à-vis de la solution exacte (l'optimum global) est lié au problème de choix qui consiste à décider s'il faut poursuivre la génération d'optima locaux, ou si au contraire il est plus opportun d'interrompre l'exploration de l'espace des solutions possibles. La formulation d'une règle itérative

- définissant le nombre k d'optima locaux qu'il faut engendrer à chaque itération pour contrôler simultanément la probabilité d'amélioration sur chacune de ces suites de k observations
- définissant des conditions d'arrêt basées sur la comparaison de la valeur espérée du gain sur les k observations de l'itération subséquente avec le potentiel d'amélioration et avec le coût total de calcul de cette itération subséquente.

constitue l'objet de la sixième section.

La septième section traite de l'utilisation d'une estimation de la solution exacte pour explorer plus efficacement l'espace des solutions possibles et ainsi engendrer une solution plus "proche" de la solution exacte. (Le qualificatif "efficace" se rapporte ici au temps d'exécution et à la place de mémoire requise). Il est proposé d'introduire cette estimation comme borne supérieure initiale dans un algorithme de "branch and bound"; on s'engage ainsi dans un algorithme heuristique à priori, parfois exact à posteriori:

ou bien la solution exacte est engendrée, ou bien on obtient une solution heuristique meilleure que celles précédemment connues, ou bien encore aucune solution n'est produite. Dans ces deux derniers cas, on déduit de nouveaux intervalles de confiance à 100% pour la solution exacte.

La première section précise l'utilité des procédures heuristiques et conséquemment justifie plus complètement les objectifs de ce papier. Les propriétés caractéristiques des techniques de hill-climbing étant à la base

- des procédures qui, pour estimer l'optimum global, font usage des optima locaux engendrés par de telles heuristiques (sections 2, 3, 4 et 5),
- de la règle qui précise les conditions d'arrêt de l'exploration réalisée par de telles heuristiques (section 6),

cette première section se poursuit par la description de ces propriétés.

1. POSITION DU PROBLEME

1.1. LA NECESSITE DES TECHNIQUES HEURISTIQUES

Dans le cas des problèmes d'optimisation combinatoire de grande taille, le nombre de solutions possibles est tellement important qu'il a fallu inventer des méthodes qui échappent à l'énumération explicite de toutes ces solutions. Ces méthodes dites "d'exploration" appartiennent à deux catégories [21,38].

a) Les méthodes exactes. Ce sont en général des procédures d'énumération implicite des solutions possibles; elles aboutissent nécessairement à la solution exacte (Branch and bound).

b) Les méthodes heuristiques. Pour certains problèmes de grande taille, les méthodes exactes s'avèrent inefficaces. Pour résoudre ces problèmes "difficiles", on a inventé des algorithmes heuristiques, c'est-à-dire des méthodes qui ne garantissent pas la génération de la solution exacte, mais qui, par contre, s'avèrent très efficaces et en général donnent de bonnes solutions. Ces procédures d'exploration heuristiques sont de deux types:

   - Les techniques d'exploration dirigée (hill-climbing), parfois couplées à des techniques de génération aléatoire (random sampling) pour produire les solutions initiales nécessaires au hill-climbing.
   - Les techniques séquentielles (constructive).

Les méthodes séquentielles construisent une solution de manière itérative en prenant à chaque itération la "décision élémentaire" la plus favorable à l'optimisation.

Les méthodes de hill-climbing démarrent avec une solution complète qu'elles

amélicrent progressivement, par petits ajustements successifs choisis selon quelques règles systématiques, et cela jusqu'à ce que plus aucune amélioration ne soit possible. La solution ainsi obtenue par convergence constitue ce qu'on appelle un optimum local [35].

Les problèmes d'optimisation combinatoire "difficiles" sont nombreux et bien connus des chercheurs opérationnels: problème du chemin le plus long, problème du voyageur de commerce, problème du sac alpin, problème d'affectation quadratique, etc... La plupart des auteurs [4,20,39,43] qui traitent de ces problèmes "difficiles" laissent entendre qu'aucun algorithme exact et efficace ne peut être inventé pour les résoudre. Plusieurs parmi ces problèmes difficiles appartiennent à ce qu'on appelle la famille "NP-complete":

> "Any current algorithm for obtaining the exact solution to a problem in this class is an exponential time algorithm. That is, in the worst case, running time increases exponentially as a function of an input parameter, such as the number of nodes" [10] (dans le cas du problème du voyageur de commerce par exemple).
> "... the concept of NP-completness (have been used) to show that these problems are equivalent in the sense that if any of them can be solved in less than exponential time, then can all" [4].
> "... mounting empirical evidence (i.e. the identification of more and more NP-complete problems) suggests that it is very likely that no polynomial time algorithms exists for such any of these problems" [39].

L'importance de ces problèmes "difficiles" et notamment de ceux qui appartiennent à la famille "NP-complete" justifie l'attention accordée aux procédures d'exploration heuristiques mentionnées plus haut. Mais si la première difficulté pour ce type de problème est l'élaboration de ces heuristiques, la seconde difficulté est celle de l'évaluation de la qualité des

solutions qu'elles produisent, c'est-à-dire l'évaluation de leur proximité vis-à-vis de la solution exacte. A quelles "distances" se trouvent-elles de la solution exacte? Comment évaluer cette distance? En résumé, on se trouve confronté avec la situation suivante:

- Il existe des problèmes d'optimisation combinatoire pour lesquels la solution exacte ne peut être obtenue dès lors qu'ils dépassent certaines tailles; des méthodes exactes ne pouvant être utilisées, on a inventé des procédures heuristiques.
- Ces heuristiques ne garantissant pas la génération de la solution exacte, il est nécessaire d'évaluer la qualité des solutions qu'elles engendrent.
- Cette qualité (les "distances") ne peut être évaluée que si la solution exacte est connue.

Par conséquent, notre objectif central sera l'estimation de cette solution exacte.

1.2. NOTATIONS.DOMAINE D'APPLICATION

L'approche développée dans ce rapport est suffisamment générale pour être appliquée à de nombreux problèmes d'optimisation combinatoire. L'ensemble de toutes les solutions possibles ρ d'un tel problème peut être représenté par M. La fonction objectif peut être représentée par X; à toute solution $\rho \in M$, cette fonction X définie sur M associe une valeur $X(\rho)$. Dans le cas d'un problème de minimisation, il faut trouver une solution ρ^* de valeur minimale, c'est-à-dire telle que

$$X(\rho^*) \leq X(\rho) \qquad \rho^* \in M; \quad \text{quel que soit} \quad \rho \in M.$$

L'auteur envisage l'expérimentation numérique de sa recherche dans le domaine de la planification urbaine et régionale. Les modèles mathématiques de planification qui utilisent les procédures d'optimisation heuristiques,

et qui par conséquent sont directement intéressés par les objectifs de ce rapport,sont récents. Il s'agit surtout de ceux qui ont pour objet la conception de plans d'utilisation des sols [16,21,23,25,26,27,29] et de plans de transport [3,21,22,24]. Certains modèles plus rares envisagent le problème de la planification intégrée de l'utilisation des sols et du réseau de transport [21,26,27,30,31]. D'autres modèles traitent spécifiquement de la localisation des services publics [28,36,40,41], de l'affectation des investissements industriels et des investissements antipollution [12,13,17,18]. Enfin, certains auteurs [15,28] considèrent la planification hiérarchisée et des services publics et des autres activités.

Pour fixer les idées, dans la suite de ce rapport on supposera avoir affaire à un problème de planification de l'utilisation des sols formulé sous la forme d'un problème d'affectation quadratique du type Koopmans-Beckmann [16]: On dispose de n activités à affecter à n localisations de manière à minimiser un coût; il existe n! solutions possibles, c'est-à-dire n! applications possibles ρ de l'ensemble des activités $A=\{1,2,\ldots,n\}$ sur l'ensemble des localisations $\mathcal{L}=\{1,2,\ldots,n\}$. Toute solution possible est une application $\rho: A \rightarrow \mathcal{L}$, application à laquelle correspond une permutation des n éléments de l'ensemble A. Ces n! applications constituent l'ensemble M. A toute permutation ρ est associée une valeur, un coût $X(\rho)$:

$$X(\rho) = \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij}\, c_{\rho(i)\rho(j)}$$

où $\rho(i)$ représente la localisation de l'activité i dans le plan ρ

q_{ij} représente l'interaction entre les activités i et j; $(i,j=1,2,\ldots,n)$

$c_{k\ell}$ représente le coût d'interaction entre les localisations k et ℓ; $(k,\ell=1,2,\ldots,n)$.

Le problème d'optimisation consiste à trouver une permutation $\rho^* \in M$ de coût minimal.

Les variables seront représentées par des lettres capitales; les valeurs prises ou réalisables seront représentées par les lettres minuscules correspondantes; on notera donc $X(\rho_i) \equiv x_i$, $X(\rho^*) \equiv x^*$, etc...

1.3. LES METHODES DE HILL-CLIMBING

L'estimation de la solution exacte peut servir à l'évaluation de n'importe quelle solution, que celle-ci provienne d'une technique de hill-climbing, d'une procédure séquentielle, ou plus simplement encore d'une méthode de génération aléatoire. C'est cependant dans le cadre des techniques de hill-climbing que l'estimation de la solution exacte s'avère la plus utile; ces techniques, en effet, rendent aisée la génération d'un grand nombre de solutions (des optima locaux): il suffit de démarrer avec différentes solutions initiales pour aboutir, en principe, à différents optima locaux; dès lors, si le meilleur optimum local atteint à un moment donné ne donne pas satisfaction compte tenu de la valeur estimée de la solution exacte (l'optimum global), il est facile de choisir de nouvelles solutions initiales et de leur appliquer l'algorithme de hill-climbing pour essayer d'engendrer de nouveaux optima locaux plus satisfaisants.

Ces techniques de hill-climbing étant, d'autre part, au centre des considérations développées dans les sections 2, 3, 4, 5 et 6, il est opportun d'en préciser ici les caractéristiques essentielles.

Toute procédure d'exploration du type hill-climbing peut être très simplement interprétée moyennant l'introduction de ce que Reiter et Sherman [35] ont appelé une "structure de succession". La formalisation qui suit s'inspire directement de leur article.

Une "structure de succession sur M" est une paire (π,s) où

a) π est une collection de sous-ensembles $\{\pi(\rho) : \rho \in M\}$ telle que

- quelque soit $\rho \in M$, il existe un sous-ensemble $\pi(\rho) \in \pi$
- quel que soit $\rho \in M$, $\rho \in \pi(\rho)$

b) s est une fonction définie sur M et telle que

- $s(\rho) \in \pi(\rho)$
- $X(s(\rho)) \leq X(\rho)$
- $X(s(\rho)) = X(\rho) \Rightarrow s(\rho) = \rho$
- $s(\rho) = \rho \Rightarrow X(\rho') \geq X(\rho)$, quel que soit $\rho' \in \pi(\rho)$.

La collection π est appelée la "structure de voisinage sur M", $\pi(\rho)$ est le "voisinage" de ρ et la fonction s est la "fonction de succession sur M". Cette structure de succession associe à chaque solution (à chaque permutation) $\rho \in M$ une solution $s(\rho)$ dans son voisinage, une solution où la fonction objectif acquiert une valeur au plus égale à $X(\rho)$. $s(\rho)$ est appelé le "successeur" de ρ.

Quelques autres définitions:

a) Si dans $\pi(\rho)$ aucune solution distincte de ρ n'a une valeur inférieure à $X(\rho)$, alors ρ est son propre successeur: $s(\rho) = \rho$. Par définition, si $\rho \in M$ est tel que $s(\rho) = \rho$, on dit que ρ est une solution localement optimale et que $X(\rho)$ est un optimum local.

b) Le $t^{\text{ième}}$ successeur de ρ, $s^t(\rho)$, est défini par $s^t(\rho) = s(s^{t-1}(\rho))$, $t = 2,3,\ldots$

c) Une solution $\rho^* \in M$ telle que $X(\rho^*) \geq X(\rho)$, quel que soit $\rho \in M$, est dite globalement optimale et $X(\rho^*)$ est un optimum global.

Ces quelques définitions décrivent ce qui se passe lors de l'utilisation d'un

algorithme d'exploration du type hill-climbing. En effet, toute heuristique de ce type consiste à partir d'une solution initiale $\rho \in M$ et à l'améliorer par petits ajustements successifs, jusqu'à ce qu'il ne soit plus possible d'améliorer la solution "courante". Ces ajustements résultent de l'application de certaines règles systématiques qui, d'une façon générale, procèdent de la manière suivante:

1. On associe à la solution "courante" le sous-ensemble de solutions qui peuvent en être déduites grâce à une légère modification dont le type est défini à priori. Ce sous-ensemble de solutions "adjacentes" à la solution courante ρ en constitue le "voisinage"; il correspond à l'un des sous-ensembles $\pi(\rho)$ définis plus haut. Dans le problème que nous avons choisi comme exemple, ρ est une permutation qui peut être représentée par $\rho = (i_1, i_2, i_3, \ldots i_n)$; si le type de modification déterminant les solutions adjacentes est défini à priori comme étant l'échange de deux activités, n'importe quelle permutation obtenue à partir de ρ grâce à un tel échange constitue une solution adjacente à ρ: ainsi, $(i_1, i_3, i_2, i_4, \ldots i_n)$ est une solution adjacente à ρ; le voisinage $\pi(\rho)$ de ρ est le sous-ensemble formé des permutations qu'on peut obtenir à partir de ρ par échange de deux activités.
2. Dans le voisinage $\pi(\rho)$ de la solution courante ρ, on recherche la solution adjacente à laquelle correspond la plus petite valeur de la fonction objectif et on implante effectivement la modification correspondant à cette "meilleure" solution adjacente; cette dernière n'est rien d'autre que le successeur $s(\rho)$ défini ci-dessus[1]; elle devient la nouvelle solution courante et le processus se poursuit jusqu'à ce qu'il ne soit plus possible d'améliorer la solution courante: on a alors at-

1. Si plusieurs solutions adjacentes sont candidates, c'est-à-dire si la fonction objectif X acquiert cette "plus petite valeur" en plusieurs points de $\pi(\rho)$, il suffit d'organiser l'algorithme pour qu'un seul des candidats soit sélectionné de telle sorte que $s(\rho)$ soit unique.

teint une solution qui est son propre successeur, on a atteint un optimum local.

Soulignons que c'est relativement à une "structure de succession" (π,s), c'est-à-dire relativement à un algorithme donné qu'une solution est dite localement optimale.

Quelques autres remarques très importantes:

a) Pour n'importe quelle "structure de succession" (π,s), il existe des solutions localement optimales. En particulier, une solution globalement optimale ρ^* est localement optimale vis-à-vis de (π,s). Cela revient aussi à dire que le sous-ensemble de solutions localement optimales relativement à (π,s) donné, c'est-à-dire relativement à un algorithme donné, contient nécessairement les solutions globalement optimales.

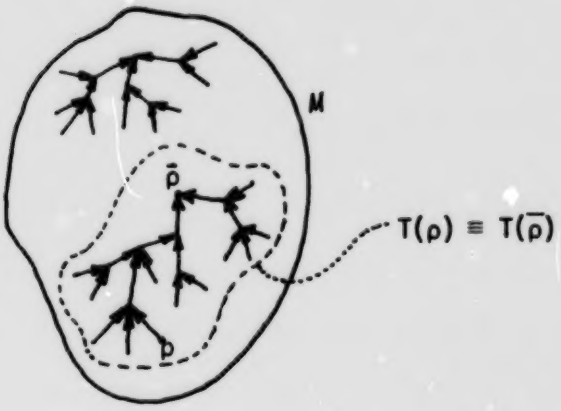
b) Pour n'importe quelle "structure de succession" (π,s) et pour toute solution $\rho \in M$, il existe un entier t (dépendant de ρ) tel que $s^t(\rho) = s^{t+1}(\rho)$; par définition, pour tout $\rho \in M$, $\bar{\rho}$ est appelé le "successeur ultime" de ρ si il existe un entier t tel que $s^{t+r}(\rho) = \bar{\rho}$, quel que soit l'entier $r \geq 0$; en d'autres termes, toute solution ρ possède un "successeur ultime" $\bar{\rho}$ qui bien sûr n'est rien d'autre qu'un optimum local; ce successeur ultime est unique. Bien sûr, des solutions distinctes ont souvent des successeurs ultimes distincts, ce qui revient à dire qu'en appliquant un algorithme donné à des solutions initiales distinctes, on aboutit souvent à des optima locaux distincts.

Une dernière définition pour conclure cette formalisation: à toute solu-

tion $\rho \in M$, on associe $T(\rho) = \{q \in M : \bar{q} = \bar{\rho}\}$; $T(\rho)$ est l'ensemble des solutions q de M qui ont le même successeur ultime que ρ. En d'autres termes $T(\rho)$ représente le sous-ensemble de M qui regroupe toutes les solutions de M qui aboutiraient au même optimum local que ρ si on leur appliquait l'algorithme. L'ensemble M peut ainsi être partitionné en sous-ensembles disjoints tels que chacun d'eux correspond à un optimal local particulier, à savoir le successeur ultime de tous les éléments de ce sous-ensemble.

En résumé, tout algorithme du type hill-climbing équivaut à la formation d'une "structure de succession sur M". Cette structure de succession provoque une partition de l'ensemble M en sous-ensembles disjoints, chacun de ceux-ci regroupant l'ensemble des solutions qui aboutiraient à un même optimum local si on les introduisait comme solutions initiales dans l'algorithme; cet optimum local est le successeur ultime commun aux éléments du sous-ensemble correspondant. A un autre algorithme, c'est-à-dire à une autre structure de succession (π, s), correspond une autre partition de M et d'autres optima locaux.

La structure de succession associée à un algorithme peut être interprétée en termes de graphe: les éléments de M sont les noeuds; un arc est orienté de ρ vers ρ' si ρ' est le successeur de ρ. Le graphe résultant consiste en un certain nombre d'arbres (inversés) disjoints, chacun d'eux possédant un optimum local en son sommet. En choisissant une solution particulière ρ comme solution initiale, la structure de succession conduit à une solution unique localement optimale, à savoir au sommet de l'arbre (unique) contenant la solution initiale ρ.



Reiter et Sherman ont également introduit la notion de "structure de probabilité":

Soit $\bar{\rho}$ le successeur ultime d'une solution quelconque ρ. Si la solution initiale est choisie au hasard, c'est-à-dire si elle est engendrée par une procédure de "génération aléatoire" (en d'autres termes si elle est sélectionnée selon une loi de probabilités uniforme), la probabilité de sélectionner une solution initiale dont le successeur ultime soit l'optimum local $\bar{\rho}$ est [1]

$$\text{prob}(\bar{\rho}) = \frac{|T(\rho)|}{|M|} \equiv \frac{|T(\bar{\rho})|}{|M|}$$

Cette probabilité est la probabilité que l'algorithme de hill-climbing combiné avec une procédure de génération aléatoire de la solution initiale conduise à l'optimum local $\bar{\rho}$. En d'autres termes, la combinaison d'un algorithme de hill-climbing et d'une procédure de génération aléatoire induit une distribution de probabilités (définie par la relation ci-dessus) sur le sous-ensemble des optima locaux associés au hill-climbing[2].

---

1. Si A est un ensemble, alors |A| représente le nombre cardinal de A.
2. En fait, n'importe quelle loi de probabilité pour sélectionner les solutions initiales induit une certaine densité de probabilité sur le sous-ensemble des optima locaux.

2. ESTIMATION DE L'OPTIMUM GLOBAL: ETAT DE L'ART

2.1. REMARQUE PRELIMINAIRE [4,20,39,43]

De nombreuses recherches ont eu pour objectif de produire des garanties sur la performance des algorithmes exacts ou heuristiques; les recherches de cette nature traitent à la fois de la complexité des problèmes et de l'efficacité des algorithmes (running time). Certaines études analysent la complexité d'un problème indépendamment de l'algorithme utilisé (lower bound on complexity: borne inférieure pour le temps de calcul nécessaire). D'autres études, regroupées sous l'expression "worst case analysis", visent à déterminer une borne supérieure pour le temps de calcul nécessaire à un algorithme donné. Enfin, certaines recherches essaient de caractériser en termes de probabilité l'ensemble des cas qui peuvent se présenter (probabilistic analysis) de façon à étudier le comportement moyen des algorithmes (average case behavior), par exemple en vue de tirer quelque information sur la valeur espérée du temps de calcul. D'une façon générale, ces recherches (theory of computational complexity) portent sur l'efficacité des algorithmes et sur l'évaluation des chances de résoudre certaines classes de problèmes dans des temps raisonnables. La famille "NP-complete" mentionnée plus haut occupe une place centrale dans ces recherches, et ces dernières ont manifestement un lien de parenté avec notre sujet; il convenait donc de les mentionner même si leurs objectifs sont parfaitement distincts de ceux qui nous occupent ici.

En ce qui concerne plus précisément les algorithmes heuristiques, des études connexes ont porté sur la qualité des solutions engendrées, mais d'un

point de vue bien différent de celui que nous avons adopté ici. D'une façon générale, ces études s'efforcent de fournir des garanties générales (a priori, c'est-à-dire indépendamment des données spécifiques caractérisant les problèmes) sur la qualité des solutions susceptibles d'être engendrées par un algorithme donné; leur but est l'évaluation, a priori, des performances minimales des algorithmes, c'est-à-dire l'évaluation de l'écart maximum (%) possible vis-à-vis de l'optimum global, quelles que soient les données des problèmes. Ces recherches sont donc également voisines du sujet de ce rapport, mais leurs objectifs en sont bien différents.

Le point de vue que nous avons adopté a pour objectif non pas la caractérisation des performances d'un algorithme heuristique, mais bien l'évaluation proprement dite de la qualité des solutions effectivement produites par cet algorithme: pour un problème donné, il s'agit d'estimer l'écart entre les solutions obtenues et l'optimum global en produisant une estimation de celui-ci.

2.2. INFERENCE STATISTIQUE CLASSIQUE ET GENERATION ALEATOIRE

Si l'on pouvait énumérer explicitement toutes les n! permutations possibles de notre problème d'optimisation combinatoire, on serait à même de calculer la probabilité qu'un élément ρ_i, choisi au hasard dans l'ensemble M regroupant toutes ces solutions, ait une valeur $X(\rho_i)$ inférieure ou égale à n'importe quelle grandeur x fixée a priori; cette probabilité serait fournie par

$$\text{Prob}\,[X(\rho_i) \leq x] = \frac{\text{nombre d'éléments de M ayant une valeur} \leq x}{|M|}$$

Ces probabilités fourniraient la fonction de répartition représentant l'ensemble M. En d'autres termes, à cette population-mère M, on peut associer une distribution statistique dès lors qu'on considère la variable X comme une variable aléatoire. Cette distribution des valeurs $X(\rho_i) \equiv x_i$ où $i = 1,2,\ldots,n!$ est malheureusement inconnue; tout ce qu'on sait, c'est qu'elle est bornée à gauche par l'optimum global [1] $X(\rho^*) \equiv x^*$ que nous cherchons.

La première idée qui vient naturellement à l'esprit est de faire appel à l'inférence statistique; en construisant des solutions au hasard et en étudiant l'échantillon ainsi constitué, il devrait être possible de tirer des conclusions au sujet de la population-mère, à condition bien sûr de pouvoir considérer que cet échantillon la représente correctement en miniature. C'est précisément un tel échantillonnage que réalise une procédure de génération aléatoire (un générateur de permutations aléatoires) puisqu'une telle procédure est équivalente à la sélection d'éléments de M selon une loi de probabilités uniforme. Si le nombre de solutions aléatoires produites est suffisamment grand, on peut considérer que la distribution de cet échantillon est une approximation de la distribution de la population-mère et que la meilleure des solutions aléatoires est une estimation de l'optimum global. Cependant, aucune hypothèse n'ayant été engagée quant à la distribution-mère, on est incapable d'évaluer l'incertitude associée à cette estimation.

Dans le même esprit, une approche très pertinente consiste à estimer la probabilité que, dans un échantillon de taille T, au moins une observation se situe dans la fraction inférieure p de la population [33]. En d'autres

---

1. Cette solution globalement optimale n'est pas nécessairement unique.

mots, on se demande quelle doit être la taille T de l'échantillon pour être certain, à 99% par exemple, qu'au moins une observation soit inférieure ou égale au quantile[1] d'ordre 0.01 (par exemple) de la distribution-mère. Un calcul élémentaire de probabilité montre qu'il faudrait constituer un échantillon de 1000 observations environ; dans des problèmes où le nombre de solutions possibles est extrêmement grand, il reste que le pourcent inférieur en question contient encore un très grand nombre de solutions, avec entre elles des écarts pouvant aussi être très importants; ceci n'est guère satisfaisant compte tenu des efforts produits pour générer les 1000 solutions aléatoires.

Quelques auteurs ont combiné les procédures de génération aléatoire avec d'autres heuristiques. M. Los [21,25,26] est arrivé à la conclusion que les procédures combinant génération aléatoire de solutions initiales et hill-climbing étaient les plus prometteuses pour les problèmes de grande taille. Mais encore une fois, n'ayant engagé aucune hypothèse statistique, les solutions obtenues ne sont accompagnées d'aucune évaluation de l'erreur dont elles sont entachées; même en produisant un grand nombre d'optima locaux et en retenant le meilleur comme approximation de l'optimum global, on ne peut fournir aucune évaluation de l'incertitude associée à cette approximation.

En fait, les premiers à avoir eu cette idée sont Reiter et Sherman [35]: "the basic idea of our approach is to combine intelligent search and random search".

Ce qu'il faut d'abord souligner, finalement, c'est que si une hypothèse de

---

1. Le quantile d'ordre p est défini comme le nombre réel qui divise l'aire sous la courbe de densité de probabilité en deux parties telles que celle de gauche représente 100p% de l'aire totale, c'est-à-dire 100p% de la population.

nature statistique n'est pas engagée quelque part, il est impossible d'inférer des conclusions garanties au sujet de la population-mère. Cette constatation n'est guère surprenante: l'inférence statistique classique (paramétrique) n'est capable d'associer une mesure de confiance aux conclusions qu'elle propose (une mesure qui soit vraiment significative) que si elle dispose d'un minimum d'information sur la population-mère ou si elle propose une hypothèse au sujet de celle-ci.

Formuler une hypothèse sur la distribution de la population-mère paraît cependant très hasardeux: a priori, cette population discrète ne répond à aucune loi, à aucune organisation intrinsèque. Pour échapper à la nécessité de formuler une hypothèse sur la distribution-mère, deux voies sont possibles; celles-ci sont présentées dans les deux paragraphes suivants; toutes deux procèdent de l'inférence statistique non-paramétrique (ou méthodes de libre distribution) [7], théorie qui a précisément pour caractéristique essentielle de rechercher des informations indépendantes de la distribution mère. La seconde voie semble la plus prometteuse; elle est davantage explorée dans les paragraphes ultérieurs.

2.3. LES PREMIERS TRAVAUX DE DANNENBRING ET CEUX DE ROBSON-WHITLOCK

La théorie des variables statistiques ordonnées constitue l'un des principaux outils de l'inférence statistique non-paramétrique; Dannenbring [2] et Robson-Whitlock [37] l'ont utilisée pour calculer un estimateur de la borne inférieure (ou supérieure) de toute distribution tronquée à gauche (ou à droite):

Si $\rho_1, \rho_2, \ldots, \rho_T$ représentent un échantillon de T permutations aléatoires, c'est-à-dire si $x_1 \equiv X(\rho_1)$, $x_2 \equiv X(\rho_2), \ldots, x_T \equiv X(\rho_T)$ constituent un échantillon de T observations tirées de la population-mère M à l'aide d'une pro-

cédure de génération aléatoire, l'estimation simple (point estimation) de la borne inférieure de cette distribution tronquée est fournie par [1]

$$\hat{X}(\rho^*) \equiv \hat{x}^* = 2x_{(1)} - x_{(2)} \qquad (1)$$

où $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(T)}$ sont les observations $x_1, x_2, \ldots, x_T$ ordonnées par ordre de valeurs croissantes.

Testé par Dannenbring sur des problèmes où l'optimum global était connu, cet estimateur a fourni des résultats moyens. En n'utilisant que les deux plus petites statistiques ordonnées $x_{(1)}$ et $x_{(2)}$, cet estimateur perd inévitablement une large partie de l'information contenue dans tout l'échantillon $x_1, x_2, \ldots x_T$. D'autre part, cet estimateur ne fait usage d'aucune solution heuristique.

De façon à utiliser l'information additionnelle que constituent les solutions qu'on a pu engendrer à l'aide de l'une ou l'autre technique heuristique, Dannenbring [2] a substitué à cet estimateur d'autres expressions qui, en fait, reprennent l'idée avancée plus haut (Reiter et Sherman) de combiner la génération aléatoire avec des procédures heuristiques. L'estimation $2x_{(1)} - x_{(2)}$ est en effet mise en défaut si l'une ou l'autre heuristique a fourni une solution de valeur inférieure à l'estimation.
Soient donc $x_{(1)} \leq x_{(2)} \leq \ldots, \leq x_{(T)}$ les valeurs ordonnées d'un échantillon de T permutations obtenues par une procédure de génération aléatoire.
Soit, d'autre part, $\hat{X}(\rho^h) \equiv x^h$ la valeur de la meilleure solution engendrée par une heuristique, quelle que soit celle-ci (hill-climbing ou technique séquentielle); cette "meilleure" solution heuristique est étrangère à l'échantillon de solutions aléatoires; Dannenbring propose les estimateurs

---

1. Le symbole ^ indiquera toujours qu'il s'agit d'estimations.

suivants:

$$\begin{aligned} \hat{x}^* \equiv \hat{X}(p^*) &= 2x_{(1)} - x_{(2)} && \text{si } x_{(2)} \leq x^h \\ &= 2x_{(1)} - x^h && \text{si } x_{(1)} \leq x^h \leq x_{(2)} \\ &= 2x^h - x_{(1)} && \text{si } x^h \leq x_{(1)} \end{aligned} \qquad (2)$$

Il propose aussi un estimateur défini comme le minimum de l'expression (1) et de l'expression (2): Min {(1),(2)}, et enfin un estimateur qu'il définit comme la moyenne arithmétique suivante: $\frac{1}{2}$ [(1)+Min {(1),(2)}]. La seule justification que Dannenbring donne à ces expressions, c'est de considérer la meilleure solution heuristique x^h comme une solution aléatoire supplémentaire. Soulignons encore une fois que ces estimateurs ignorent l'information contenue dans le reste de l'échantillon $x_1, x_2, \ldots, x_T$, ainsi que celle contenue dans les autres solutions heuristiques disponibles. Si par exemple, on possède un algorithme de hill-climbing, il est facile d'engendrer un grand nombre de solutions heuristiques: les estimateurs proposés ci-dessus n'envisagent l'utilisation que du meilleur de ces optima locaux.

Robson et Whitlock [37] sont allés plus loin en produisant un intervalle de confiance à 100(1-α)%. Ceci constitue un progrès significatif vis-à-vis des résultats précédents parce que l'obtention d'un intervalle de confiance est un problème beaucoup plus compliqué que celui de l'estimation simple. Malheureusement, encore une fois, cet intervalle n'utilise que les deux plus petites statistiques ordonnées de l'échantillon de solutions aléatoires: il perd toute l'information contenue dans les (T-2) observations $x_{(3)}, x_{(4)}, \ldots, x_{(T)}$, et d'autre part il ne fait aucun usage des solutions que l'une ou l'autre heuristique a engendré. Robson et Whitlock ont montré que leur

intervalle de confiance, exact pour une distribution-mère uniforme, devient approximatif pour les autres distributions:

$$\text{Prob}\{x_{(1)} - (\frac{1-\alpha}{\alpha})(x_{(2)} - x_{(1)}) \leq x^* \leq x_{(1)}\} \approx 1 - \alpha \qquad (3)$$

A noter que pour $\alpha = 0.5$, la limite inférieure s'identifie avec l'estimateur de l'expression (1).

2.4. LES PREMIERES RECHERCHES INSPIREES DE FISHER-TIPPETT

Lorsqu'on dispose d'une procédure de hill-climbing, il est facile d'engendrer de nombreuses solutions heuristiques (voir 1.3). Si $\rho_1^h, \rho_2^h, \ldots, \rho_N^h$ sont N permutations tirées d'une telle procédure h, ces N solutions, en tant qu'optima locaux, doivent nécessairement contenir une information beaucoup plus riche que les solutions engendrées à l'aide d'une procédure de génération aléatoire ou que les quelques solutions provenant d'une technique séquentielle; tous les estimateurs proposés ci-dessus ne tirent aucun parti de cette possibilité. L'idée, dans ce qui suit, est de proposer une hypothèse non pas sur la distribution de la population-mère, mais plutôt sur la distribution des optima locaux engendrés par une procédure de hill-climbing: si on pouvait formuler une hypothèse sur la distribution des optima locaux, on devrait être à même, à partir d'un échantillon de ces optima locaux, de tirer quelque information de nature probabiliste sur l'optimum global (puisqu'il n'est qu'un optimum local particulier), ou encore sur les chances d'amélioration du meilleur optimum local atteint si l'exploration heuristique était poursuivie. Les premiers à avoir imaginé cette approche sont Reiter et Sherman [35]: ce sont eux (voir 1.3) qui ont montré que toute procédure de hill-climbing combinée avec

l'une ou l'autre loi de probabilités pour sélectionner les solutions initiales induit une densité de probabilité bien spécifique sur l'ensemble des optima locaux associés à l'algorithme de hill-climbing envisagé. Malheureusement, Reiter et Sherman n'avaient aucune idée sur la forme de cette densité de probabilités ; dans l'impossibilité de proposer telle hypothèse plutôt que telle autre, la procédure qu'ils ont imaginée est restée dans le domaine de la pure heuristique.

L'approche la plus prometteuse, et c'est sur elle que se fonde l'essentiel de la démarche développée dans la suite de ce rapport, est due à McRoberts [33]. Cette approche poursuit l'idée suggérée ci-dessus de faire porter l'attention sur la distribution des optima locaux[1]; elle s'inspire de la théorie des valeurs extrêmes (théorie au sujet de laquelle on trouve la plus complète information dans Gumbel [11]), et plus précisément d'un théorème énoncé pour la première fois en 1928 par Fisher et Tippett [5].

Considérons N échantillons indépendants, chacun de taille m, issus d'une population-mère de densité f(X). Dans chaque échantillon, il existe une observation de valeur minimale x_i (i=1,...,N), de sorte que finalement on dispose d'un "super-échantillon" $x_1, x_2, \ldots x_N$ de valeurs extrémales. En fait, chaque observation minimale est la valeur prise par la plus petite variable statistique ordonnée d'un échantillon de taille m, variable dont la distribution dépend de m et de f(X). Lorsque la taille m augmente, la distribution de cette variable "extrémale" approche une distribution

---

1. Par conséquent, dans la suite du rapport, lorsqu'on parlera de procédures, de techniques... heuristiques, il ne s'agira plus que des procédures du type hill-climbing. Toutes les solutions heuristiques envisagées seront donc des "optima locaux".

limite. Toutes les distributions initiales f(X) ne possèdent pas une telle distribution asymptotique (voir conditions ci-dessous), mais lorsque cette distribution asymptotique existe, Fisher et Tippett [5] ont montré qu'elle ne peut être que l'un ou l'autre parmi trois types de base (types I, II et III de Gumbel). Dans le cas où l'on s'intéresse à la distribution asymptotique de la valeur minimale correspondant à une distribution-mère f(X) bornée à gauche, la distribution asymptotique est de type III, à savoir ce qu'il est convenu d'appeler aujourd'hui la distribution de Weibull [42], ou plus exactement la famille Weibull puisqu'il s'agit d'une distribution contenant trois paramètres. En d'autres termes, lorsque la taille m de chacun des N échantillons indépendants issus d'une distribution quelconque f(X) bornée à gauche devient très grande, la distribution des valeurs minimales $x_i (i=1,\ldots,N)$ de ces N échantillons approche une distribution connue, indépendante de la distribution-mère f(X), et dite "distribution de Weibull":

fonction de répartition $\Phi(X) = 1 - \exp\left[-\left(\frac{X-a}{b}\right)^c\right]$

densité de probabilité $\delta(X) = \frac{c}{b}\left(\frac{x-a}{b}\right)^{c-1} \exp\left[-\left(\frac{X-a}{b}\right)^c\right]$

$X \geq a \geq 0; c>0; b>0$

Les paramètres a, b et c sont respectivement les paramètres de localisation, d'échelle et de forme.

Soulignons que, normalement, l'existence d'une asymptote suppose que la distribution-mère est continue; c'est Gnedenko [8] qui a mis en évidence les conditions nécessaires et suffisantes à l'existence d'une distribution asymptotique de type III pour la valeur minimale; ces conditions, reprises dans [32], peuvent s'écrire:

- L'étendue sur laquelle est définie la distribution-mère[1] F(X) doit être bornée inférieurement: $F(X) \equiv 0$ pour $X \leq a$

1. F(X) représente la fonction de répartition.

-F(X) doit se comporter comme $b(X-a)^c$ pour $b,c>0$ lorsque $X \to a$.

Pour ce qui nous occupe ici, l'intérêt fondamental de la théorie résulte des deux constatations suivantes:

- La distribution des valeurs minimales est indépendante de la distribution de la population-mère.
- Le paramètre de localisation a de la loi Weibull n'est rien d'autre que la borne inférieure de la distribution-mère.

McRoberts [33] fut le premier à appliquer cette théorie pour déduire une estimation simple (point estimation) du paramètre de localisation comme approximation de l'optimum global. L'idée fondamentale est la suivante: la loi Weibull étant une fonction mathématique décrivant le comportement de valeurs extrêmes (observées dans des échantillons indépendants tirés de n'importe quelle population-mère bornée à gauche), on peut imaginer que la distribution des optima locaux engendrés par une procédure heuristique approche cette loi Weibull et cela quelle que soit la distribution de la population-mère; par conséquent, l'estimation du paramètre de localisation de cette loi fournira une approximation de l'optimum global (borne inférieure de la distribution): on échappe ainsi à la nécessité de formuler une hypothèse au sujet de la distribution-mère. (On retombe dans le domaine de l'inférence statistique non-paramétrique). L'argumentation de McRoberts est cependant affaiblie en substance par le fait qu'il considéra que chacun des échantillons de taille m du théorème de Fisher-Tippett était formé de la succession[1] des solutions construites à l'aide du hill-climbing, à partir d'une solution initiale (aléatoire) jusqu'à un optimum local: de toute évidence, l'indépendance qui doit caractériser les éléments formant un

---

1. Dans le cadre des notations du paragraphe 1.3, une telle suite peut être représentée par $\rho \to s^1(\rho) \to s^2(\rho) = s(s^1(\rho)) \to \ldots \to s^{t+1}(\rho) = s(s^t(\rho)) = s^t(\rho) = \bar{\rho}$.

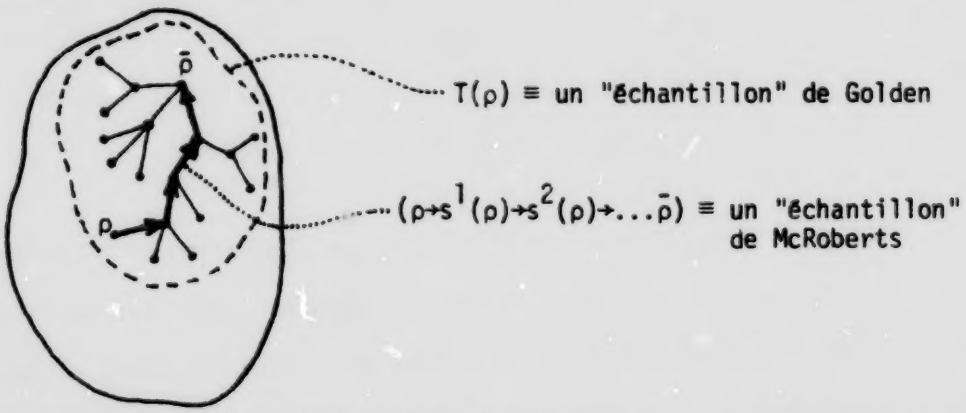
échantillon n'était aucunement vérifiée. D'autre part, McRoberts ne donna aucune indication quant à la précision de ses estimations.

Dannenbring [2] a appliqué la même théorie des valeurs extrêmes directement aux solutions produites par une procédure de génération aléatoire. En générant des échantillons de solutions aléatoires (il est allé jusqu'à 40 000 solutions), il a appliqué l'hypothèse Weibull à la distribution des extrema de ces échantillons. En d'autres termes, il a construit des estimateurs asymptotiques utilisant des extrema d'échantillons aléatoires véritablement tirés de la population-mère. Ces estimateurs (étrangers à la notion d'optimum local) se sont avérés moins précis que ceux qu'il élabora sur base des travaux de Robson-Whitlock (voir paragraphe 2.3) et qui eux faisaient usage de la plus petite solution heuristique x^h. Bref, Dannenbring a profité de l'analogie suggérée par McRoberts quant à l'utilité de la distribution asymptotique de valeurs extrêmes, mais il s'est écarté de l'idée selon laquelle c'est la distribution des optima locaux liés à une heuristique qui est la clef du problème.

Clough [1], utilisant les techniques de simulation du type Monte-Carlo, a appliqué la théorie des valeurs extrêmes pour estimer l'optimum global d'une fonction de plusieurs variables soumises a des contraintes. Il a décrit les distributions d'échantillonnage de ses estimateurs, ce qui lui a permis de déduire des intervalles de confiance. Malheureusement, Clough s'est limité au cas où le paramètre de forme c=1, se ramenant ainsi au cas particulier d'une distribution exponentielle.

2.5. LES RECHERCHES DE GOLDEN ET ALT

Golden et Alt [9,10] ont repris l'ensemble des travaux précédents en se penchant plus particulièrement sur le problème du voyageur de commerce. Dans un premier article, Golden [9] a utilisé la loi Weibull pour obtenir d'excellentes estimations simples (point estimation) de l'optimum global; l'idée générale est semblable à celle de McRoberts, mais l'hypothèse de base est quelque peu corrigée: pour Golden, ce n'est plus la succession des solutions construites durant l'application d'une procédure de hill-climbing à une solution (aléatoire) initiale jusqu'à un optimum local qui constitue l'un des N échantillons de taille m (Fisher-Tippett); chacun de ces échantillons est plutôt constitué du sous-ensemble particulier et implicite de solutions possibles qui aboutiraient à un même optimum local si on leur appliquait le hill-climbing; ces sous-ensembles implicites ont été notés T(ρ) dans le paragraphe 1.3:



En d'autres termes, chaque "échantillon" de McRoberts est inclu dans l'un des "échantillons" de Golden; plus précisément, il correspond à un chemin particulier dans l'arbre qui peut représenter l'un des "échantillons" de Golden. Mis à part cet ajustement de l'hypothèse centrale, le principe de la méthode est inchangé: chaque optimum local est considéré comme la

valeur minimale d'un échantillon implicite; la théorie des valeurs extrêmes suggère alors l'hypothèse résultante selon laquelle la distribution des optima locaux résultant de l'application de l'heuristique à des solutions initiales aléatoires approche la loi Weibull.

L'estimation simple des trois paramètres de la distribution Weibull, à partir d'un échantillon de taille N (N optima locaux), est un problème relativement compliqué. La littérature propose différentes méthodes. McRoberts [33] a utilisé la méthode des moindres carrés. Gumbell [11] défend l'usage de la méthode des moments. Dannenbring [2] conseille la méthode des moindres carrés plutôt que celle des moments. En fait, lorsque les trois paramètres sont inconnus, les procédures d'estimation devraient être itératives [32] et utiliser le principe du maximum de vraisemblance. C'est ce principe qu'utilise la procédure de Harter-Moore [14] qui est recommandée par différents auteurs, notamment par Mann, Schafer et Singpurwalla [32] et surtout par Zanakis [44] qui a procédé à la comparaison de plusieurs méthodes. Golden a obtenu d'assez bons résultats avec la méthode des moindres carrés, mais dans un second article [10] écrit en collaboration avec Alt, il utilise la technique de Harter-Moore.
Soulignons aussi que pour réaliser de telles estimations à partir d'un "échantillon" de N optima locaux, une autre condition doit être satisfaite: les N observations doivent être indépendantes.

L'essentiel du premier article de Golden est non seulement d'avoir donné une évaluation de la qualité de ses estimations [1], mais surtout d'avoir tiré parti de toutes les possibilités offertes par la théorie de l'inférence statistique pour vérifier les hypothèses:

---

1. Golden a travaillé sur des problèmes dont l'optimum global était connu.

Pour vérifier l'hypothèse selon laquelle la distribution des optima locaux approche la loi Weibull, Golden a utilisé le test de Kolmogorov-Smirnov (voir annexe 1). Ce test d'ajustement compare la distribution théorique (hypothétique) avec la distribution empirique. Tous les tests furent positifs.

Pour vérifier l'hypothèse d'indépendance des observations (les N optima locaux), Golden a utilisé des tests qui vérifient le caractère aléatoire de suites d'observations (voir annexe 2). Ces tests furent également positifs.

En résumé, le schéma méthodologique général est le suivant:

a) Hypothèse centrale: chaque optimum local peut être interprété comme la valeur extrémale d'un échantillon implicite.

b) La théorie des valeurs extrêmes suggère l'hypothèse résultante selon laquelle la distribution des optima locaux approche la loi Weibull (conséquence testable).

c) La théorie de l'inférence statistique est alors utilisée
   - pour estimer (à partir d'un échantillon de N optima locaux supposés indépendants) la distribution Weibull et ainsi inférer une estimation de la borne inférieure de la population-mère;
   - pour vérifier les hypothèses.

Il reste que de telles estimations simples ne fournissent aucune idée, aucune évaluation de l'erreur qu'elles contiennent; les seules garanties sont fournies indirectement par les tests de vérification des hypothèses. L'estimation d'un paramètre ne devient vraiment significative que si elle précise la probabilité d'erreur qu'elle comporte. Il faudrait donc

a) déterminer pour x^* un intervalle de variation de la forme $[\hat{x}^* - \beta, \hat{x}^* + \gamma]$

b) évaluer la probabilité que x* se trouve dans cet intervalle.

La production d'un intervalle de confiance constitue l'objet du papier de Golden et Alt [10].

Le premier intervalle qu'ils ont testé s'inspire de celui fourni par Robson et Whitlock (voir l'expression (3) en page 21).

Golden et Alt ont appliqué cet intervalle à la distribution des optima locaux. L'expression (3) de cet intervalle approximatif devient donc

$$\text{Prob}\ \{x^h_{(1)} - (\tfrac{1-\alpha}{\alpha})(x^h_{(2)} - x^h_{(1)}) \leq x^* \leq x^h_{(1)}\} \simeq 1 - \alpha \qquad (4)$$

où $x^h_{(1)}$ et $x^h_{(2)}$ sont respectivement la plus petite et la seconde plus petite observations correspondant à un échantillon $x^h_1 \equiv X(\rho^h_1)$, $x^h_2 \equiv X(\rho^h_2)$,... ..., $x^h_N \equiv X(\rho^h_N)$ de N solutions heuristiques (N optima locaux).

La faiblesse de cet intervalle est qu'il est indépendant de la distribution des optima locaux et donc qu'il n'exploite aucunement l'idée selon laquelle cette distribution pourrait être la loi Weibull.

Golden et Alt ont finalement produit un intervalle de confiance axé sur la meilleure solution heuristique $x^h_{(1)}$, et cela en tirant profit d'une propriété à laquelle la loi Weibull obéit toujours:

$$\Phi(a+b) = 1 - e^{-1} \simeq 0.63 \qquad (5)$$

Les N observations x_i^h (i=1,...N) forment, par hypothèse, un échantillon issu de la distribution Weibull; il est donc possible d'écrire:

$$\text{Prob}\{x_{(1)}^h \leq a + b\}$$

$$= 1-\text{Prob}\{x_{(1)}^h > a + b\}$$

$$= 1-\{1-\Phi_{X_1^h}(a+b)\}\{1-\Phi_{X_2^h}(a+b)\}\ldots\{1-\Phi_{X_N^h}(a+b)\}$$

$$= 1-(e^{-1})^N = 1-e^{-N}$$

$$\Rightarrow \quad \text{Prob}\{x_{(1)}^h - b \leq a \leq x_{(1)}^h\} = 1-e^{-N}$$

En introduisant dans cette expression l'estimation $\hat{b}$ du paramètre b, on obtient un intervalle de confiance approximatif à $100(1-e^{-N})\%$ pour a, c'est-à-dire pour x*

$$\text{Prob}\{x_{(1)}^h - \hat{b} \leq x^* \leq x_{(1)}^h\} \simeq 1-e^{-N} \qquad (6)$$

Notons que cet intervalle n'est pas axé sur l'estimation simple $\hat{x}^* \equiv \hat{a}$, mais sur la meilleure solution heuristique $x_{(1)}^h$. Cet intervalle a remarquablement fonctionné dans toutes les expériences numériques réalisées par Golden et Alt.

Le paramètre b fut estimé [1] à l'aide de la technique de Harter-Moore initialisée à l'aide des expressions suivantes [2] :

---

1. La procédure de Harter-Moore fournit une estimation des trois paramètres.
2. $x_{(1)}^h \leq x_{(2)}^h \leq \ldots \leq x_{(N)}^h$ sont les observations $x_1^h, x_2^h, \ldots, x_N^h$ ordonnées par ordre de valeurs croissantes.

$$\hat{a} = 2\, x^{h}_{(1)} - x^{h}_{(2)} \qquad (7)$$

$$\hat{b} = x^{h}_{([0.63N]+1)} - \hat{a} \qquad \text{où } [0.63N] \text{ est le plus grand entier inférieur à } 0.63N \qquad (8)$$

$$\hat{c} = \frac{\ell n(-\ell n(0.5))}{\ell n(x^{N}_{M}-\hat{a}) - \ell n\hat{b}} \qquad \text{où } x^{N}_{M} \text{ est la médiane des N observations} \qquad (9)$$

Les équations (7) et (8) proviennent respectivement des équations (4) et (5). L'équation (9) est obtenue en prenant deux fois le logarithme de la fonction de répartition $\Phi(X)$ de la loi Weibull et en remplaçant X par X^{N}_{M}, puis a,b, et c par leurs estimateurs.

2.6. CRITIQUE DE GOLDEN ET ALT

Ce paragraphe souligne les faiblesses les plus manifestes de l'approche développée par Golden et Alt. C'est cependant la section 3 qui se réserve la discussion de la nuance la plus délicate qu'il convient d'apporter aux considérations théoriques sous-jacentes de l'approche en question. Parallèlement à ces réserves quant aux fondements proprements dits de ladite approche, c'est d'autre part la qualité même de l'intervalle de confiance défini ci-dessus qui doit faire l'objet d'une critique non moins importante; cette dernière question est traitée au début de la quatrième section. La présentation est ainsi divisée pour de simples raisons de clarté.

a) En fait, l'hypothèse selon laquelle les optima locaux ont une distribution qui approche la loi Weibull est le résultat d'une hypothèse préalable. Chaque optimum local associé à un algorithme de hill-climbing est effectivement la valeur minimale d'un sous-ensemble de solutions, à savoir le sous-ensemble regroupant toutes les solutions possibles qui aboutiraient à l'optimum local en question si on leur appliquait l'heuristique; l'hypothèse intervient lorsqu'on déclare que chacun de ces optima locaux se comporte comme la valeur extrême d'un échantillon aléatoire. En d'autres termes, on suppose l'indépendance des solutions qui constituent implicitement un échantillon;

cette hypothèse est extrêmement discutable puisque, par définition, des solutions ont précisément en commun la caractéristique d'aboutir au même optimum local si le hill-climbing leur est appliqué: c'est donc qu'elles partagent une propriété essentielle qui les lie les unes aux autres. Les "échantillons" de McRoberts sont inclus dans les "échantillons" de Golden et Alt, et l'intuition laisse penser que l'indépendance de leurs éléments est davantage garantie dans le cas de Golden et Alt; il n'en reste pas moins vrai que cette indépendance ne peut être certifiée, bien au contraire comme le montre la remarque précédente.

En résumé, l'hypothèse selon laquelle les optima locaux constituent implicitement des valeurs extrêmes au sens de la théorie statistique ne peut être rigoureusement justifiée, encore que tout dépend du concept même d'indépendance, vocable auquel on peut prêter différentes interprétations (indépendance fonctionnelle, indépendance au sens de probabilités...)

b) Le concept d'indépendance est également mis en cause aux niveaux des observations (les N optima locaux engendrés). Dans leur cas, les tests de caractère aléatoire (appliqués à de telles suites d'optima locaux obtenus à partir de solutions initiales aléatoires) ont convaincu Golden et Alt de leur indépendance (pour ces tests, voir l'annexe 2). Ces optima locaux ne sont cependant pas "étrangers" les uns aux autres puisqu'ils dépendent de l'algorithme utilisé (voir le paragraphe 1.3) et bien sûr, de la structure empirique du problème (les données proprement dites). En d'autres termes, les optima locaux liés à une heuristique et à un problème sont "heuristiquement" et "empiriquement" dépendants les uns des autres. Toutefois, encore ici, c'est le concept même d'indépendance qui requiert quelque éclaircissement.

c) L'un des aspects les plus discutables est delui de la pertinence même de la démarche du point de vue de l'approximation d'une distribution discrête par une distribution continue. Nous savons que cette approximation est fausse: la variable X ne peut prendre qu'un nombre fini de valeurs possibles (n! dans le cas du problème que nous avons choisi pour exemple), et qui plus est, les valeurs extrémales (les optimà locaux) sont en nombre encore plus limité.

En réalité, et ceci a été souligné précédemment, l'existence d'une distribution asymptotique suppose que la distribution-mère est continue et que la taille (m) des échantillons est grande. Dans le cas des problèmes combinatoires, les distributions sont discrètes et rien ne garantit l'existence même d'une asymptote. Tout ce qu'on peut espérer finalement, c'est que le phénomène d'explosion combinatoire autorise l'approximation qui est effectuée, mais il reste que la preuve théorique rigoureuse de la pertinence de l'hypothèse Weibull est impossible. Tout ce qui est réalisé, c'est une analogie avec la théorie des valeurs extrêmes. Seule l'expérience pourra donc appuyer ou au contraire rejeter cet emprunt. *D'une façon générale d'ailleurs, la preuve que les considérations qui sous-tendent l'hypothèse Weibull (continuité, indépendance...) ne sont pas acceptables est du ressort de l'expérience.*

Remarque: Que la taille de certains échantillons implicites ne soit pas nécessairement grande (alors que le théorème de Fisher-Tippet [5] suppose au contraire que m soit grand) est, semble-t-il, un inconvénient majeur. Dannenbring [2] signale cependant avoir obtenu les meilleures estimations asymptotiques lorsque, pour un nombre total donné d'observations (Nm), la taille m des échantillons était petite mais leur nombre (N) était grand. Cette analyse de la sensibilité des estimations aux nombres N et m fut rendue possible par le

fait que Dannenbring (voir paragraphe 2.4) s'intéressait à des échantillons directement et effectivement extraits de la population-mère (et non pas à des optima locaux et à leurs sous-ensembles associés). Le résultat de Dannenbring est très intéressant en ce sens qu'il laisse penser que la taille même des échantillons implicites associés aux N optima locaux engendrés jouerait un rôle moins important que le nombre N.

d) Un autre problème existe au niveau de la méthode de vérification de l'hypothèse Weibull. Golden et Alt ont utilisé le test d'ajustement Kolmogorov-Smirnov. Tous les tests furent positifs, ce qui les a conduits à confirmer l'hypothèse. En fait, les tests furent sans doute exagérément positifs (la distribution normale, qui est une approximation de la loi Weibull, a également répondu positivement), ce qui laisse penser que le test n'est peut-être pas assez puissant.

A côté du test de Kolmogorov-Smirnov, les statisticiens utilisent souvent un autre test d'ajustement qui est celui du χ^2. Celui-ci exige cependant que les données soient groupées en classes numériques. En ce sens, le test de Kolmogorov-Smirnov utilise plus complètement l'information puisqu'il teste l'ajustement en toutes les observations, alors que le test du χ^2 ne vérifie cet ajustement qu'en un nombre réduit de points. D'autre part, la distribution d'échantillonnage de la statistique (le critère de test) du test de K-S est connue et tabulée alors que celle du test du χ^2 n'est qu'approximativement "chi-carré".

Le problème qui se pose avec ces tests est que les paramètres de la distribution théorique sont (dans notre cas) inconnus et doivent être estimés à partir des données, lesquelles précisément forment la distri-

bution empirique: il en résulte un test qui n'est qu'approximatif [6,7]. C'est sans doute la raison essentielle pour laquelle on peut craindre que ces tests ne soient pas assez puissants. De ce point de vue, lorsque les paramètres doivent être estimés, l'avantage du test du χ^2 est qu'il peut être accompli de la façon habituelle (c'est-à-dire comme dans le cas où les paramètres sont connus); la seule différence est qu'il faut réduire le nombre de degrés de liberté d'un nombre égal au nombre de paramètres estimés.

D'autre part, les propriétés théoriques du test de K-S requièrent qu'on ait affaire à une distribution continue, ce qui n'est pas le cas avec la distribution que nous traitons. Différents articles [6] ont cependant montré que le test de K-S réalisé comme si la distribution était continue reste correct, à condition de remplacer les équations de probabilité par des inéquations.

Signalons enfin qu'un papier de Lilliefors mentionné par Gibbons [7] permet de penser que le test de K-S réalisé après estimation des paramètres est plus puissant que le test du χ^2.

3. DISTRIBUTIONS PROPREMENT DITE ET INDUITE

Résumons brièvement l'approche développée jusqu'ici. Reiter et Sherman ont montré que la combinaison d'un algorithme de hill-climbing et d'une procédure de génération aléatoire des solutions initiales induit une distribution de probabilités sur le sous-ensemble (désignons-le par H) des optima locaux associés à l'algorithme utilisé.

S'inspirant de la théorie des valeurs extrêmes, Golden et Alt en arrivent

à proposer l'hypothèse selon laquelle cette distribution des optima devrait approcher la loi Weibull: cette hypothèse résulte de l'application du théorème de Fisher-Tippett aux optima locaux interprétés comme les valeurs extrêmes d'échantillons implicites.

En réalité, cette analogie avec la théorie des valeurs extrêmes n'est pas rigoureusement correcte: lorsqu'on estime la loi Weibull supposée à partir d'un échantillon de N optima locaux et que dans cet échantillon on retrouve plusieurs fois les mêmes optima, cette façon de procéder équivaut en fait à utiliser plusieurs fois les mêmes échantillons implicites, alors que le théorème de Fisher-Tippett requiert au contraire l'indépendance des N échantillons. Dans nos premiers essais numériques (non repris dans ce rapport) sur le problème d'affectation quadratique, nous nous sommes ainsi retrouvés dans un cas [1] où, parmi les N=100 optima de l'échantillon destiné à estimer la loi Weibull, le plus petit apparaissait 24 fois; il en résultait d'une part une valeur du paramètre de localisation égale à ce plus petit optimum local et d'autre part la valeur 0.24 pour la fonction de répartition empirique en ce point: inéluctablement, l'écart vertical maximum entre la distribution Weibull théorique et la distribution empirique, écart qui constitue le critère du test de Kolmogorov-Smirnov, s'est situé au droit de cette plus petite valeur observée et cet écart (égal à 0.24) important a conduit ainsi au rejet de l'hypothèse.

Là où une erreur théorique est commise, c'est en prétendant que la distribution de probabilités induite par l'association du hill-climbing et de

---

1 Il s'agissait d'un problème de taille 11 : 11 activités à affecter à 11 localisations.

la procédure de génération aléatoire des solutions initiales — distribution définie par Reiter et Sherman — doit approcher la loi Weibull. La distribution empirique (sample or empirical cumulative distribution function) des N optima locaux x_1^h, x_2^h, ..., x_N^h est effectivement une approximation de la loi de probabilités décrivant le comportement du hill-climbing associé à la procédure de génération aléatoire des solutions initiales, c'est-à-dire de la loi de probabilités décrivant la distribution des optima susceptibles de résulter de ladite association, mais on ne peut faire l'hypothèse que cette distribution "induite" approche la loi Weibull parce qu'il est incorrect d'appliquer ici le théorème de Fisher-Tippet: chaque fois que l'un des optima se répète p fois dans l'échantillon de taille N, cela revient à utiliser p fois le même échantillon implicite au moment de l'application "analogique" du théorème en question.

Par contre, si dans l'échantillon de taille N on élimine toutes les répétitions pour ne conserver qu'un échantillon de $R \leq N$ optima locaux <u>distincts</u> on supprime du même coup les répétitions des mêmes échantillons implicites et rien n'empêche alors d'appliquer le théorème de Fisher-Tippet à ces R optima distincts:
Soient donc x_1^{hd}, x_2^{hd}, ..., x_R^{hd} les $R \leq N$ optima distincts de l'échantillon original x_1^h, x_2^h, ..., x_N^h des N optima locaux résultants de l'application du hill-climbing à N solutions initiales produites de façon aléatoire. Comme précédemment, on peut interpréter chaque observation x_i^{hd} (i=1,2,...,R) comme la valeur extrême d'un échantillon implicite. On dispose donc de R valeurs extrêmes "extraites" de R échantillons distincts "issus" de la même population-mère, R extrema à propos desquels l'hypothèse Weibull semble

cette fois justifiée[1]. L'intervalle de confiance défini par Golden et Alt (expression (6)) devient

$$\text{Prob}\{x^{hd}_{(1)} - \hat{b} \leq x^* \leq x^{hd}_{(1)}\} \simeq 1 - e^{-R} \qquad (10)$$

où $\hat{b}$ est l'estimation du paramètre d'échelle résultant (par exemple) de l'application de la procédure de Harter-Moore aux R observations distinctes, procédure initialisée à l'aide des expressions (7),(8) et (9) qui s'écrivent cette fois

$$\hat{a} = 2x^{hd}_{(1)} - x^{hd}_{(2)} \qquad (11)$$

$$\hat{b} = x^{hd}_{([0.63R]+1)} - \hat{a} \qquad \text{où } [0.63R] \text{ est le plus grand entier inférieur à } 0.63R \qquad (12)$$

$$\hat{c} = \frac{\ell n(-\ell n(0.5))}{\ell n(x^R_M - \hat{a}) - \ell n\hat{b}} \qquad \text{où } x^R_M \text{ est la médiane des R observations} \qquad (13)$$

$x^{hd}_{(1)} < x^{hd}_{(2)} < \ldots < x^{hd}_{(R)}$ étant les observations distinctes $x^{hd}_1, x^{hd}_2, \ldots x^{hd}_R$ ordonnées par ordre de valeurs croissantes.

En fait, en ne conservant que les R optima distincts, tout se passe comme si on tirait directement un échantillon de taille R de la population *H* constituée de tous les optima locaux associés à l'algorithme de hill-climbing utilisé. La distribution empirique de ces R optima est une approximation de la distribution statistique que l'on pourrait former si on pouvait énumérer explicitement tous les optima locaux (tous les éléments du sous-ensemble *H*). En d'autres termes, c'est à propos de la distribution de cette sous-population *H* (distribution que nous appelerons "la distribution proprement dite" des optima locaux) que l'on peut faire l'hypothèse Weibull en donnant à chaque élément de *H* la signification d'une

1. Les considérations restrictives du paragraphe 2.6 continuent bien sûr à s'appliquer.

valeur extrême au sens du théorème de Fisher-Tippett, et non pas à propos de la loi de probabilités décrivant la distribution des optima susceptibles d'être engendrés à la suite de l'application du hill-climbing à des solutions initiales aléatoires, distribution "induite" dont parlaient Reiter et Sherman.

Cette distinction importante n'a pas été signalée par McRoberts, Golden et Alt parce que, vraisemblablement, dans leurs essais numériques, tous les optima engendrés, ou presque, étaient distincts les uns des autres, de telle sorte que les distributions qu'ils ont estimées furent effectivement les distributions "proprement dites" en question. De fait, lorsqu'on a affaire à des problèmes combinatoires de grande taille, la plupart des optima qu'on engendre, même s'ils sont nombreux, sont presque tous différents: la distinction entre la distribution qu'approchent les N optima originaux (la distribution "induite" de Reiter et Sherman) et la distribution qu'approchent les R optima distincs (la distribution "proprement dite") est négligeable. Dans le cas d'un problème d'affectation quadratique de taille 24 (24 activités à affecter à 24 localisations) par exemple, sur un échantillon de 140 optima, 137 étaient distincts les uns des autres; pour un problème de taille 48, toujours sur 140 optima, nous n'avons jamais rencontré deux fois le même optimum local. En d'autres mots, pour les problèmes de grande taille, la distinction théorique introduite ci-dessus devient dérisoire, et on peut considérer que la distribution "proprement dite" des optima estimée à partir d'un échantillon de R ($\simeq$N) optima distincts, est quasi identique à la distribution des optima induite par le hill-climbing et la procédure de génération aléatoire des

solutions initiales: ces deux distributions devraient suivre la même loi Weibull. Il reste bien sûr la question de savoir à partir de quel moment (à partir de quelle taille minimale des problèmes) on peut admettre cette équivalence, c'est-à-dire de savoir quand on peut considérer que l'estimation de la distribution "proprement dite" des optima (à partir des R optima distincts) est en même temps une estimation de la distribution "induite", une estimation de la loi des probabilités décrivant le comportement du hill-climbing et de la procédure de génération aléatoire des solutions initiales. Cette question n'est pas purement académique: elle est d'une importance capitale pour ce qui est de la règle d'arrêt, qui fait l'objet de la section 6, règle pour laquelle c'est précisément la distribution "induite" qui joue un rôle central.

La réponse à cette question sera fournie en étudiant des problèmes de tailles diverses, et pour chacun d'eux, en vérifiant l'ajustement de la distribution empirique des N optima locaux originaux (les observations avant suppression des répétitions) à la distribution Weibull [1] "proprement dite" estimée à partir des R optima distincts. Si le test d'ajustement de ces N optima est positif [2], on pourra étendre l'hypothèse Weibull à la distribution "induite": on disposera donc à la fois de l'estimation de la distribution "proprement dite" des optima nécessaire aux estimations simples et intervalles de confiance pour l'optimum global et d'une

---

1. Moyennant bien sûr la vérification préalable de l'ajustement des R optima distincts à la forme Weibull.

2. Ceci implique que R≈N, laquelle situation ne se présente que dans le cas des problèmes de grande taille, c'est-à-dire le cas où les estimations de la loi Weibull à partir des N optima originaux et à partir des R optima distincts conduisent quasiment aux mêmes résultats.

estimation de la loi de probabilités décrivant le comportement du hill-climbing et de la procédure de génération aléatoire des solutions initiales, loi qui est essentielle à la règle d'arrêt dont il sera question plus loin.

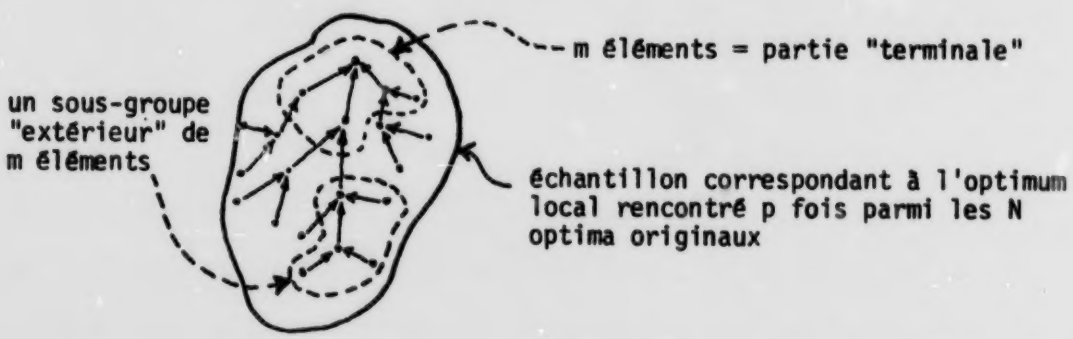
Remarquons que la distinction qui vient d'être établie entre les distributions "induite" et "proprement dite" des optima locaux résulte du caractère discret des problèmes traités. Pour les petites tailles, dans un échantillon initial de N optima, on retrouve souvent les mêmes observations; la distinction est significative et l'hypothèse Weibull n'est justifiable que pour la distribution "proprement dite", distribution qui est estimée à partir des R optima distincts. Pour les grandes tailles, les N optima engendrés sont presque tous distincts et on peut considérer que l'estimation de la distribution Weibull "proprement dite" est aussi une estimation de la distribution"induite": le phénomène d'explosion combinatoire auquel on assiste a pour effet

- d'augmenter considérablement le nombre de solutions possibles
- d'augmenter considérablement le nombre d'optima locaux susceptibles d'être engendrés
- de réduire de façon significative la probabilité d'engendrer deux fois le même optimum local.

En d'autres mots, le phénomène d'explosion combinatoire nous déplace du côté des situations où l'on aurait affaire à des distributions de type continu, des situations où toutes les argumentations théoriques développées jusqu'ici — notamment l'argumentation de l'hypothèse Weibull — sont largement plus fondées.

Une dernière remarque pour conclure cette discussion.

Si parmi les N optima locaux engendrés on retrouve par exemple p fois le même optimum, cela signifie que l'échantillon implicite qui lui correspond est plus grand [1] (intuitivement p fois plus peuplé) que ceux qui correspondent aux autres optima qu'on ne rencontre qu'une seule fois parmi les N observations. En éliminant p-1 fois l'observation répétée pour ne conserver que les optima distincts, on élimine effectivement le vice de procédure qui consistait à utiliser p fois le même échantillon implicite au moment de l'application du théorème de Fisher-Tippet; il reste cependant que l'échantillon implicite correspondant à l'optimum local qu'on a conservé est plus grand que les autres échantillons implicites; or, le théorème de Fisher-Tippett requiert l'utilisation d'échantillons de même taille (m). En d'autres termes, en appliquant le théorème en question, tout se passe comme si on procédait à une troncature de l'échantillon implicite trop grand, tout se passe comme si on ne prenait en considération que la partie "terminale" de l'arbre (cf. paragraphe 1.3) qui peut représenter cet échantillon, une partie contenant m éléments.



1. Remarquons aussi que si cet optimum local obtenu p fois est le plus petit parmi les N qu'on a engendrés, rien cependant ne permet de croire que l'on ait affaire à l'optimum global.

Les éléments de l'échantillon en question qui sont "extérieurs" à la partie terminale de l'arbre "retenue" pour pouvoir appliquer le théorème de Fisher-Tippett pourraient en fait être partitionnés en autres sous-groupes de taille m; en ne retenant que le sous-groupe terminal, on élimine du même coup les autres sous-groupes (d'autres échantillons de taille m). Il n'y aurait guère de problème théorique si les sous-groupes éliminés étaient sélectionnés au hasard, mais tel n'est pas le cas: les valeurs extrêmes de chacun de ces sous-groupes sont nécessairement toutes supérieures à la valeur extrême (l'optimum local rencontré p fois) du sous-groupe qui est conservé. En résumé, en ne conservant qu'une fois l'optimum local rencontré p fois, on évite correctement l'utilisation de p fois le même échantillon implicite, mais il reste que la troncature également implicite de l'échantillon implicite associé à l'optimum conservé introduit un biais dans l'application du théorème de Fisher-Tippett: un certain nombre d'autres échantillons de taille m (les sous-groupes "extérieurs") sont écartés, et ceux-ci ne sont pas quelconques puisque leurs propres valeurs extrêmes sont toutes nécessairement plus grandes que celle de la partie "terminale" implicitement utilisée comme échantillon de taille m dans l'application du théorème de Fisher-Tippett. Remarquons cependant encore une fois que ce biais est négligeable dans le cas des problèmes combinatoires de grande taille puisque, pour ceux-ci, s'il y a des répétitions parmi les N optima originaux, celles-ci sont rares, ce qui laisse penser que même si les sous-ensembles implicites correspondant à ces optima répétés sont plus peuplés que ceux associés aux optima "uniques", la différence de taille ne doit pas être importante. D'autre part, il reste à voir dans quelle mesure l'hypothèse d'une taille identique (m)

pour les échantillons du théorème de Fisher-Tippett ne pourrait pas être relaxée dans le cas des problèmes combinatoires de grande taille.

Si une conclusion générale peut être tirée de cette discussion très intuitive, c'est en soulignant le vide théorique qui reste à combler entre les problèmes de nature "très" discrète et les problèmes de type continu. Le phénomène d'explosion combinatoire auquel on assiste dans le cas des problèmes de grande taille suggère de multiples approximations dont les limites et les conséquences n'ont, à notre connaissance, jamais été explorées.

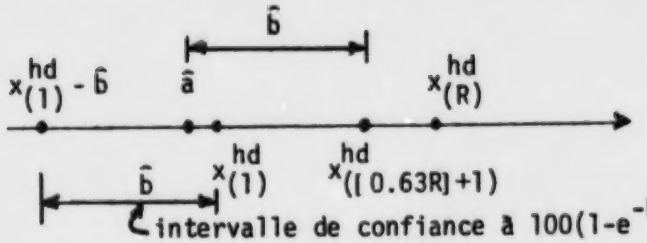
4. PROLONGEMENTS

Cette section se propose de prolonger l'approche introduite dans les pages précédentes de trois façons. La première définit un nouvel intervalle de confiance de qualité supérieure à celui de Golden et Alt. La seconde suggère l'utilisation d'heuristiques différentes. La troisième fait porter l'attention sur les valeurs extrémales d'échantillons d'optima locaux.

4.1. INTERVALLE DE CONFIANCE

L'intervalle de confiance défini par Golden et Alt (voir expression (10)) est très peu significatif: même si sa largeur ($\tilde{b}$) est relativement petite par rapport à la plus petite observation $x^{hd}_{(1)}$ ou par rapport à l'estimation $\hat{a}$ de l'optimum global, il reste que cette largeur est très grande par rapport à la grandeur $x^{hd}_{(R)} - x^{hd}_{(1)}$ qui représente l'étendue de variation des R optima distincts; en première approximation, cette largeur est

$$\hat{b} = x^{hd}_{([0.63R]+1)} - \hat{a}$$



Par exemple, dans le cas d'un problème d'affectation quadratique de taille 24, un échantillon de 50 optima distincts a donné les résultats suivants (procédure de Harter-Moore pour les estimations $\hat{a}$, $\hat{b}$ et $\hat{c}$):

$$\left.\begin{aligned} x^{hd}_{(1)} &= 142068247 \\ x^{hd}_{(50)} &= 144928416 \end{aligned}\right\} \rightarrow \quad x^{hd}_{(50)} - x^{hd}_{(1)} = 2860169$$

$$\hat{a} = 141691918 = x^{hd}_{(1)} - 0.00265 x^{hd}_{(1)}$$

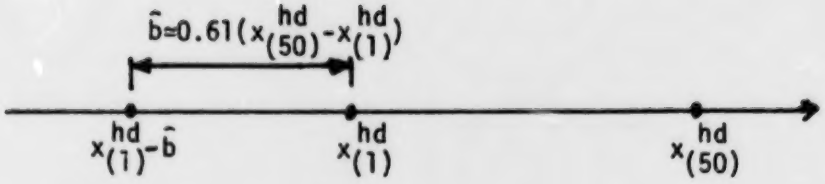
$$\hat{b} = 1751508 \simeq 0.61 \left(x^{hd}_{(50)} - x^{hd}_{(1)}\right)$$

$$\hat{c} = 2.8773$$

$$\text{Prob}\left\{x^{hd}_{(1)} - \hat{b} \leq x^* \leq x^{hd}_{(1)}\right\} \simeq 1 - e^{-50}$$

$$\text{Prob}\{140316\,739 \leq x^* \leq 142068247\} \simeq 1$$

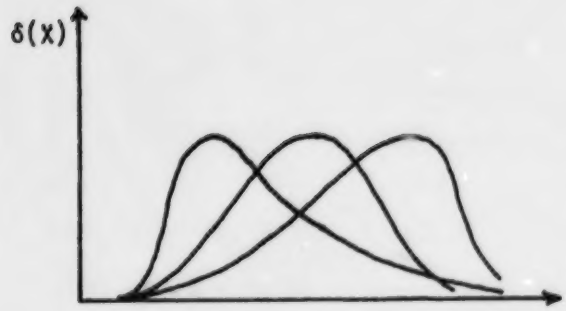
En d'autres termes, après avoir obtenu 50 observations $x^{hd}_{(1)} < x^{hd}_{(2)} < \ldots < x^{hd}_{(50)}$, tout ce que cet intervalle exprime, c'est que la vraie valeur du paramètre de localisation a (l'optimum global) doit être, avec un degré de confiance quasi égal à 100%, au moins égal à une valeur $x^{hd}_{(1)} - \hat{b}$ inférieure à la plus petite observation d'une quantité représentant 61% de l'étendue de variation des observations.



Il est évident que ce résultat est trivial: l'intervalle en question n'exprime rien qui ne soit évident et ce sera toujours le cas puisque la largeur $\hat{b}$ de cet intervalle sera toujours grande relativement à l'étendue de variation $x^{hd}_{(R)} - x^{hd}_{(1)}$. Le rapport $\hat{b}/x^{hd}_{(1)}$ que Golden et Alt utilisent comme mesure de performance de leur intervalle n'a en réalité aucunement la signification d'une telle mesure. Ce qui est significatif, c'est le rapport de $\hat{b}$ à l'une ou l'autre mesure de dispersion des observations.

En fait, la faiblesse de l'intervalle de Golden et Alt résulte essentiellement de l'indépendance de ses limites inférieures et supérieures par rapport à la probabilité $(1-e^{-R})$ que ces limites encadrent la vraie valeur du paramètre de localisation: il n'y a aucune relation de dépendance entre la largeur $\hat{b}$ de l'intervalle et le degré de confiance qui lui est associé. <u>En d'autres termes, il ne s'agit pas d'un véritable intervalle de confiance au sens de la théorie statistique</u>.

Remarquons d'autre part que la non intervention du paramètre de forme dans l'intervalle de Golden et Alt est très regrettable; ce paramètre joue un rôle non négligeable dans l'étalement de la partie gauche de la courbe qui représente la densité de probabilité; à un paramètre de forme de plus grande valeur correspond un "étalement" plus prononcé de cette partie gauche: pour un même degré de confiance, il devrait donc lui correspondre un intervalle de confiance plus large.



En suivant sensiblement le même cheminement que Golden et Alt, il est heureusement possible d'aboutir à un intervalle de confiance qui soit véritablement significatif. C'est ce qui est réalisé ci-dessous.

Soit S un nombre réel quelconque. La fonction de répartition d'une loi Weibull doit toujours satisfaire la relation suivante:

$$\Phi(a+\frac{b}{S}) = 1 - \exp\,[-(\frac{1}{S})^c]$$

Les optima locaux distincts $x_i^{hd}(i=1,\ldots,R)$ d'un échantillon de taille R doivent donc satisfaire

$$\text{Prob}\,\{x_{(1)}^{hd} \le a+\frac{b}{S}\} = 1 - \text{Prob}\,\{x_{(1)}^{hd} > a+\frac{b}{S}\}$$

$$= 1 - \{1 - \Phi_{X_1^{hd}}(a+\frac{b}{S})\}\,\{1 - \Phi_{X_2^{hd}}(a+\frac{b}{S})\}\ldots\{1 - \Phi_{X_R^{hd}}(a+\frac{b}{S})\}$$

$$= 1 - [\exp(-1/S^c)]^R = 1 - e^{-R/S^c}$$

$$\Rightarrow \text{Prob}\,\{x_{(1)}^{hd} - \frac{b}{S} \le a \le x_{(1)}^{hd}\} = 1 - e^{-R/S^c}$$

En introduisant dans cette expression les estimations $\hat{b}$ et $\hat{c}$ des paramètres d'échelle et de forme, on obtient un intervalle de confiance approximatif pour le paramètre a (c'est-à-dire pour x*), intervalle dont la largeur $\hat{b}/S$ et le degré de confiance sont cette fois directement reliés l'un à l'autre par l'intermédiaire du nombre S:

$$\text{Prob}\ \{x^{hd}_{(1)} - \frac{\hat{b}}{S} \leq x^* \leq x^{hd}_{(1)}\} \simeq 1 - e^{-R/S^{\hat{c}}}$$

Remarquons d'autre part que le degré de confiance, et donc la largeur de l'intervalle, sont cette fois fonction du paramètre forme.

Si on pose $1 - e^{-R/S^{\hat{c}}} = 1 - \alpha$, la relation entre le degré de confiance $1 - \alpha$ et - le nombre S qui intervient dans la largeur de l'intervalle

- le paramètre de forme c
- le nombre d'observations R

s'écrit plus simplement

$$S = \left(-\frac{R}{\ell n \alpha}\right)^{1/\hat{c}}$$

En se fixant a priori un degré de confiance $1 - \alpha$, et après avoir estimé les paramètres b et c à partir des R observations (procédure de Harter-Moore par exemple), on en déduit immédiatement la valeur de S et donc la largeur de l'intervalle correspondant à ce degré de confiance.
Pour des valeurs données de R, $\hat{b}$ et $\hat{c}$, une augmentation du degré de confiance $1 - \alpha$ est suivie à juste titre d'une augmentation de la largeur de l'intervalle.
De même, pour des valeurs de R et α fixées, des distributions de même $\hat{b}$ mais dont les valeurs de $\hat{c}$ sont croissantes conduisent à des intervalles de confiance plus larges; cette constatation est correcte: à des courbes plus étalées à gauche correspondent des intervalles de moindre "qualité".

Dans le cas de l'exemple cité au début de ce paragraphe, l'intervalle de confiance à 90% est donné par

$$S = \left(- \frac{50}{\log(0.10)} \right)^{1/2.8773} = 2.914649$$

$$\text{Prob } \{142068247 - \frac{1751508}{2.914649} \leq x^* \leq 142068247\} \simeq 0.90$$

$$\text{Prob } \{141467314 \leq x^* \leq 142068297\} \simeq 0.90$$

La largeur de cet intervalle à 90% ne représente plus que 21% de l'étendue de variation des observations.

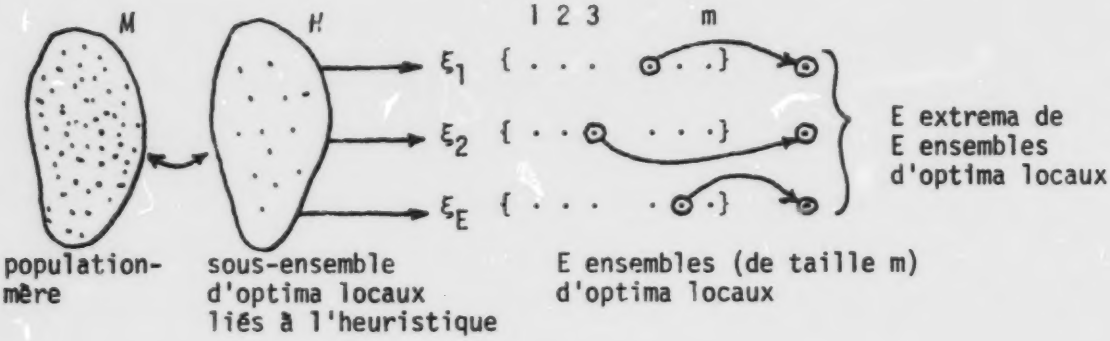
4.2. UTILISATION D'HEURISTIQUES DIFFERENTES

Les $R \leq N$ observations distinctes qui servent à l'estimation de la distribution "proprement dite" des optima locaux sont "heuristiquement" dépendantes les unes des autres (voir paragraphes 1.3 et 2.6): la "structure de succession" propre à l'algorithme a pour conséquence que ce sont telles solutions et non telles autres qui deviennent des optima locaux. En réalité, ce que requiert le théorème de Fisher-Tippet, c'est que les observations dont on identifie la distribution avec la loi Weibull soient les valeurs extrémales d'échantillons indépendants (suffisamment grands); rien n'est spécifié quant à la façon avec laquelle ces observations sont obtenues; ce qui importe, c'est de pouvoir interpréter les observations comme les extrema d'échantillons "issus" d'une même population.
En conséquence, au lieu de n'utiliser qu'une heuristique pour générer les R optima locaux distincts, c'est au contraire plusieurs heuristiques qui peuvent intervenir. Cette façon de procéder aurait l'avantage

- de produire des observations parmi lesquelles certaines ne seraient plus "heuristiquement" dépendantes;
- de multiplier le nombre d'observations pouvant jouer un rôle puisque l'ensemble des optima locaux susceptibles d'apparaître serait cette fois constitué de la réunion des ensembles d'optima locaux respectivement liés aux différentes heuristiques.

4.3. PASSAGE A UN NIVEAU "SUPERIEUR"

Les éléments constituant l'un des "échantillons" de Golden et Alt ne sont pas indépendants (voir paragraphes 1.3 et 2.6): ils partagent la propriété d'aboutir au même optimum local. Cette propriété, à la fois gênante et implicite, conduit à l'idée de réaliser véritablement la production d'échantillons, et dans chacun d'eux d'y extraire la valeur minimale: on génère E ensembles de taille m, chacune des m observations de chacun des E ensembles étant une solution heuristique; ensuite, dans chacun des E ensembles, on prend la valeur minimale.



Soient donc E ensembles ξ_i (i=1,...,E) de m optima locaux

$$\xi_i = \{x_1^{h,i}, x_2^{h,i}, \ldots, x_m^{h,i}\} \qquad (i=1,\ldots,E)$$

Soient $x_{(1)}^{h,i}, x_{(2)}^{h,i}, \ldots, x_{(m)}^{h,i}$ les observations de l'ensemble ξ_i ordonnées par ordre de valeurs croissantes[1].

Dans la mesure où les E ensembles ξ_i peuvent être considérés comme des échantillons indépendants, on aboutit finalement à un "super-échantillon"[2]

---

1. Ici, rien n'interdit que dans un même ensemble ξ_i on retrouve plusieurs fois la même observation.
2. Il se peut aussi que le même extremum $x_{(1)}^{h,i}$ se retrouve plusieurs fois dans le "super-échantillon"

$$x_{(1)}^{h,1},\ x_{(1)}^{h,2},\ \ldots,\ x_{(1)}^{h,E}$$

dont chaque élément $x_{(1)}^{h,i}$ (i=1,...,E) est la valeur extrémale d'un échantillon (de taille m) issu du sous-ensemble *H* des optima locaux liés à l'heuristique. Le théorème de Fisher-Tippett [5] suggère la nouvelle proposition suivante: la distribution des extrema $x_{(1)}^{h,i}$(i=1,...,E) devrait approcher la loi Weibull; l'estimation de son paramètre de localisation, à partir des E observations $x_{(1)}^{h,i}$(i=1,...,E) fournira une nouvelle aproximation de l'optimum global.

Ce niveau théorique est "supérieur" à ce qui fut décrit précédemment parce que:

a) Les échantillons ne sont plus issus de la population-mère M, mais bien du sous-ensemble *H* des optima locaux. (Au niveau précédent, chaque sous-ensemble dont on "tirait" la valeur extrémale était un sous-ensemble "issu" de *M*); les Em observations traitées sont donc plus "proches" de l'optimum global.

b) Les échantillons ne sont plus implicites: chaque échantillon dont on extrait la valeur extrémale est véritablement engendré.

c) Les éléments constitutifs d'un tel échantillon sont des optima locaux (en principe générés à partir de solutions aléatoires indépendantes); l'indépendance de ces m observations est discutable (elles sont "heuristiquement" et "empiriquement" liées les une aux autres - Voir paragraphe 2.6), mais elle peut être testée. Les tests de caractère aléatoire devront donc être appliqués à chacun des groupes ξ_i de m optima locaux.

Ces tests devront également être réalisés sur le "super-échantillon" des E extrema $h_{(1)}^{h,i}(i=1,\ldots,E)$.

Il va sans dire que le test d'ajustement de la distribution empirique des E extrema $x_{(1)}^{h,i}(i=1,\ldots,E)$ avec la loi théorique (Weibull) à laquelle on identifie la distribution de ces E extrema devra aussi être réalisé.

5. RECAPITULATION ET EVALUATION DES ESTIMATEURS

Ce qui suit schématise le programme de travail pour expérimenter l'approche développée dans les pages précédentes. Dans une première phase, la taille n des problèmes traités devra être suffisamment petite de façon à pouvoir comparer, en dernière instance, les différentes estimations avec la solution exacte effective. Selon les résultats fournis par les tests de vérification des hypothèses et par cette ultime évaluation, on disposera, ou non, d'une technique qui pourra dans une seconde phase être étendue à des problèmes de grande taille, des problèmes dont la solution exacte est inconnue. (Bien sûr, les tests de vérification des hypothèses devront encore être appliqués à ces problèmes). Cette extension à des problèmes de grande taille est d'autant plus justifiée que c'est précisément pour ces grandes tailles que toutes les considérations théoriques à la base de notre approche sont les mieux fondées.

5.1. ESTIMATIONS POUR UN PROBLEME DONNE

1) Niveau 1

Engendrer N permutations aléatoires $\rho_1,\rho_2,\ldots,\rho_N$ et les évaluer.

Construire l'estimateur suivant (Robson-Whitlock; Dannenbring)

$$\hat{x}_1^* = 2x_{(1)} - x_{(2)}$$

où $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(N)}$ sont les observations $x_1 \equiv X(\rho_1)$, $x_2 \equiv X(\rho_2)$, ..., $x_N \equiv X(\rho_N)$ ordonnées par ordre de valeurs croissantes.

2) Niveau 2

Engendrer N optima locaux x_1^h, x_2^h, ..., x_N^h à l'aide d'un algorithme de hill-climbing. On utilise pour solutions initiales aléatoires les permutations engendrées au niveau précédent.

Soit x^h = la meilleure solution heuristique

$$= \text{Min } \{x_i^h | i=1,\ldots,N\}$$

$$= x_{(1)}^h$$

$x_{(1)}^h \leq x_{(2)}^h \leq \ldots \leq x_{(N)}^h$ étant les observations x_1^h, x_2^h, ..., x_N^h ordonnées par ordre de valeurs croissantes.

Construire les estimateurs suivants[1] (Dannenbring):

$$\hat{x}_{2a}^* = 2x^h - x_{(1)}$$

$$\hat{x}_{2b}^* = \text{Min } \{\hat{x}_1^*, \hat{x}_{2a}^*\}$$

$$\hat{x}_{2c}^* = \frac{1}{2}(\hat{x}_1^* + \hat{x}_{2b}^*)$$

Déterminer l'intervalle de confiance de Robson-Whitlock à l'aide de la relation (4)

3) Niveau 3

Soient x_1^{hd}, x_2^{hd}, ..., x_R^{hd} les $R \leq N$ optima distincts de l'échantillon de taille N obtenu au niveau précédent. A partir de ces R optima distincts,

---

1. Ayant utilisé les solutions aléatoires x_1, x_2, ..., x_N comme solutions initiales, on a certainement $x^h \leq x_{(1)}$.

réaliser une estimation simple (point estimation) des trois paramètres de la loi Weibull à laquelle on identifie, par hypothèse, la distribution "proprement dite" des optima locaux associés à l'algorithme utilisé. Pour réaliser cette estimation, on utilise par exemple la procédure de Harter-Moore initialisée à l'aide des expressions (11), (12) et (13).

Réaliser ensuite les tests (annexes 1 et 2)

- de Kolmogorov-Smirnov pour tester l'ajustement des R observations à la distribution Weibull hypothétique,
- de caractère aléatoire pour vérifier l'hypothèse d'indépendance des R observations.

Selon les résultats des tests, on dispose ou non d'un estimateur asymptotique $\hat{x}_3^*$.

Choisir un degré de confiance $1-\alpha$ pour l'intervalle de confiance et déterminer celui-ci à l'aide des relations

$$S = \left(-\frac{R}{\ln \alpha}\right)^{1/\hat{c}} \tag{14}$$

$$\text{Prob}\left\{x_{(1)}^{hd} - \frac{\hat{b}}{S} \leq x^* \leq x_{(1)}^{hd}\right\} \simeq 1 - \alpha \tag{15}$$

4) Niveau 4

Il s'agit de réaliser une étude absolument analogue à celle du niveau précédent, mais en utilisant plusieurs algorithmes pour générer les R optima locaux distincts. On peut, par exemple, combiner un algorithme qui réalise des échanges deux à deux des activités avec un algorithme qui réalise des échanges trois à trois.

Selon les résultats des tests, on dispose ou non d'un estimateur

asymptotique $\hat{x}_4^*$.

Déterminer un ou des intervalles de confiance à partir des relations (14) et (15).

5) Niveau 5

Générer E ensembles $\xi_i = \{x_1^{h,i}, x_2^{h,i}, \ldots, x_m^{h,i}\}$ de m optima locaux.

Soient $x_{(1)}^{h,i} = \text{Min } \{x_j^{h,i} | j=1,\ldots,m\}$ $\quad i=1,\ldots,E$

A partir du "super-échantillon $(x_{(1)}^{h,1}, x_{(1)}^{h,2}, \ldots, x_{(1)}^{h,E})$ des E extrema, procéder à une estimation simple des trois paramètres de la loi Weibull à laquelle on identifie, par hypothèse, la distribution des extrema en question. Les relations (7), (8) et (9) pour initialiser la procédure de Harter-Moore deviennent

$$\hat{a} = 2x_{(1)}^{h,(1)} - x_{(1)}^{h,(2)}$$

$$\hat{b} = x_{(1)}^{h,([0.63E]+1)} - \hat{a} \qquad \text{où } [0.63E] \text{ est le plus grand entier inférieur à } 0.63E$$

$$\hat{c} = \frac{\ell n(-\ell n(0.5))}{\ell n(x_M^E - \hat{a}) - \ell n\, \hat{b}} \qquad \text{où } x_M^E \text{ est la médiane des E extrema } x_{(1)}^{h,1},\ldots,x_{(1)}^{h,E}$$

$x_{(1)}^{h,(1)} \leq x_{(1)}^{h,(2)} \leq \ldots \leq x_{(1)}^{h,(E)}$ étant les observations $x_{(1)}^{h,1}, x_{(1)}^{h,2},\ldots,x_{(1)}^{h,E}$ ordonnées par ordre de valeurs croissantes.

Réaliser le test de Kolmogorov-Smirnov pour tester l'ajustement des E extrema à la loi Weibull hypothétique.

Les tests de vérification de l'indépendance des observations doivent également être appliqués; il s'agit cette fois de vérifier le carac-

tère aléatoire de chaque suite $x_1^{h,i},\ldots x_m^{h,i}$ (i=1,...,E), ou tout au moins de la suite d'extrema $x_{(1)}^{h,1}$, $x_{(1)}^{h,2},\ldots,x_{(1)}^{h,E}$.

Selon les résultats de ces tests, on dispose ou non d'un nouvel estimateur asymptotique $\hat{x}_5^*$.

Déterminer ensuite un ou des intervalles de confiance à l'aide des expressions (14) et (15) qui deviennent

$$S = \left(-\frac{E}{\ell n\, \alpha}\right)^{1/\hat{c}}$$

$$\text{Prob}\left\{x_{(1)}^{h,(1)} - \frac{\hat{b}}{S} \leq x^* \leq x_{(1)}^{h,(1)}\right\} \simeq 1 - \alpha$$

6) Niveau 6

Il s'agit de réaliser une étude absolument analogue à celle du niveau précédent, mais en utilisant différents algorithmes pour générer les E m optima locaux.

Selon les résultats des tests, on déduit ou non un estimateur asymptotique $\hat{x}_6^*$.

Il reste à dériver un ou des intervalles de confiance en utilisant les relations énoncées ci-dessus.

5.2. COMPARAISON ET EVALUATION DES ESTIMATEURS

La qualité de l'une ou de l'autre des procédures d'estimation est indirectement évaluée par la cohérence même des considérations de nature théorique et par la vérification (les tests) des hypothèses impliquées dans cette procédure. Il reste que ces différentes procédures doivent être évaluées et comparées quant à leur précision et leur coût budgétaire. Ces

évaluations et comparaisons doivent être réalisées non seulement sur l'un ou l'autre problème, mais surtout du point de vue de la performance moyenne des procédures sur plusieurs problèmes.

L'évaluation "directe" de la précision des estimations n'est possible que pour les problèmes de petite taille puisque seuls ceux-ci peuvent être traités à l'aide d'une méthode d'optimisation exacte. Pour cette évaluation "directe", on définit [2] la mesure de précision suivante:

$$\text{Erreur relative} = 100\ (\hat{x}^* - x^*) / x^*$$

La figure No 1 schématise la mise en place des méthodes, théories, observations et tests pour la construction et l'évaluation de l'estimateur asymptotique fondé sur l'estimation de la distribution Weibull de R optima locaux distincts (estimateurs $\hat{x}_3^*$ et $\hat{x}_4^*$); un schéma analogue pourrait résumer la démarche pour l'estimateur asymptotique fondé sur l'estimation de la distribution Weibull des extrema de E échantillons de m optima locaux (estimateurs $\hat{x}_5^*$ et $\hat{x}_6^*$).

En ce qui concerne l'évaluation des intervalles de confiance, il convient d'examiner

- le rapport de la largeur de l'intervalle à l'une ou l'autre mesure de dispersion des observations, par exemple leur étendue ou leur variance
- le nombre de cas où l'intervalle contient, ou non, la solution exacte.

REMARQUE

L'expérimentation de l'approche développée dans les pages précédentes devrait se terminer par une analyse de la sensibilité des estimateurs

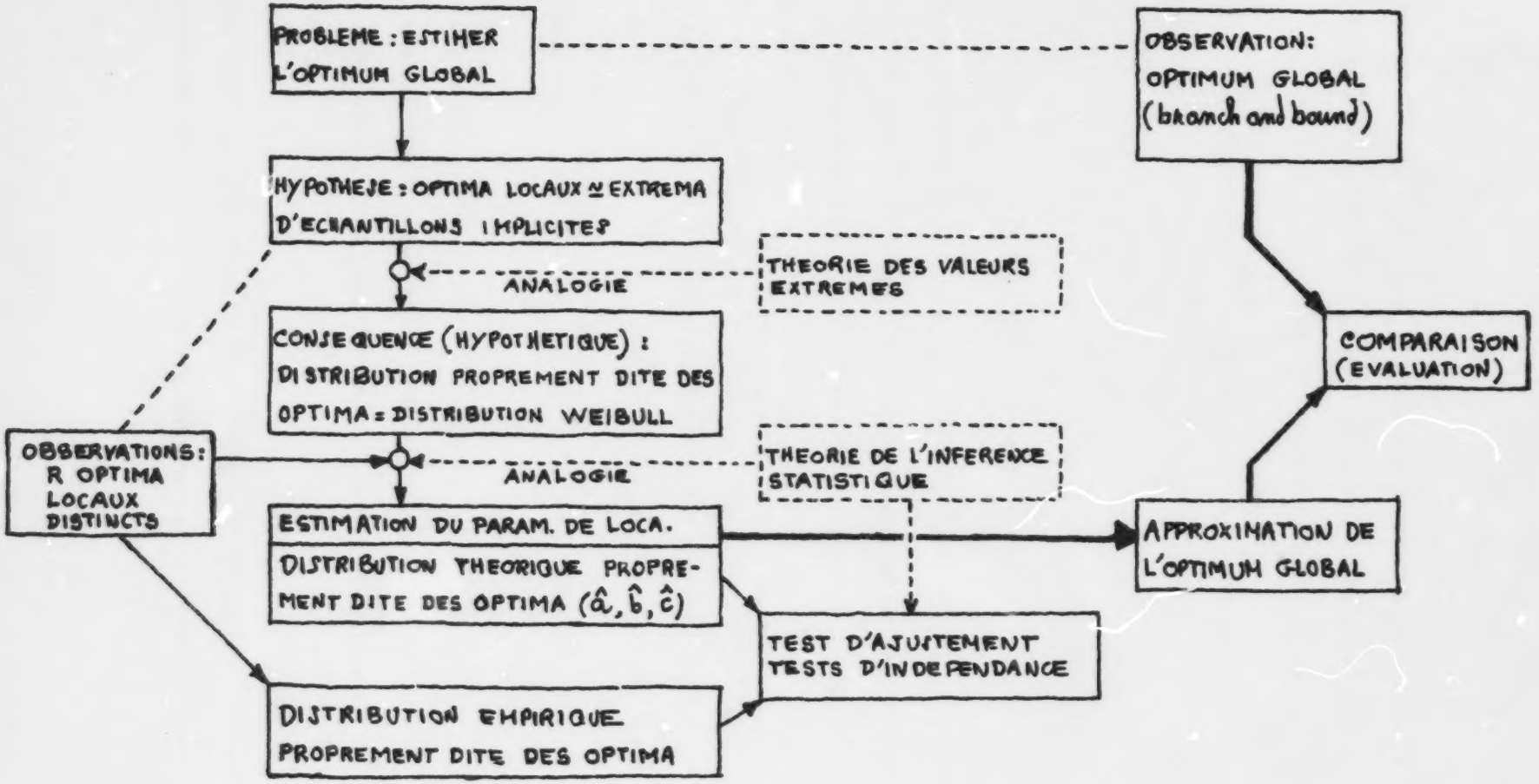

PROBLEME : ESTIMER L'OPTIMUM GLOBAL
OBSERVATION: OPTIMUM GLOBAL (branch and bound)
HYPOTHESE : OPTIMA LOCAUX ≃ EXTREMA D'ECHANTILLONS IMPLICITES
ANALOGIE
THEORIE DES VALEURS EXTREMES
CONSEQUENCE (HYPOTHETIQUE) : DISTRIBUTION PROPREMENT DITE DES OPTIMA = DISTRIBUTION WEIBULL
COMPARAISON (EVALUATION)
OBSERVATIONS: R OPTIMA LOCAUX DISTINCTS
ANALOGIE
THEORIE DE L'INFERENCE STATISTIQUE
ESTIMATION DU PARAM. DE LOCA.
DISTRIBUTION THEORIQUE PROPREMENT DITE DES OPTIMA ($\hat{a}, \hat{b}, \hat{c}$)
APPROXIMATION DE L'OPTIMUM GLOBAL
TEST D'AJUSTEMENT TESTS D'INDEPENDANCE
DISTRIBUTION EMPIRIQUE PROPREMENT DITE DES OPTIMA

FIGURE No. 1

- à la taille n des problèmes,
- aux algorithmes eux-mêmes.

Cette analyse devrait apporter quelque élément de réponse aux questions suivantes:

- A partir de quelle taille telle ou telle procédure d'estimation devient-elle justifiée?
- Pour les procédures fondées sur l'hypothèse Weibull (estimateurs asymptotiques), que se passe-t-il lorsqu'on a affaire à de mauvais algorithmes? En particulier, comment s'ajustement les observations à la courbe théorique, notamment dans la partie gauche de celle-ci?

6. REGLE D'ARRET

Lorsqu'on utilise une procédure de hill-climbing pour résoudre un problème d'optimisation combinatoire de grande taille, la première difficulté est celle de l'évaluation de la qualité des optima locaux générés. Cette question a été traitée dans les pages précédentes.

Le second problème que posent ces heuristiques est de nature plus pratique: il s'agit de leur utilisation "économique", c'est-à-dire de pouvoir déterminer si, après avoir généré un certain nombre N d'optima locaux et après avoir procédé à l'estimation de l'optimum global (estimation réalisée à partir des $R \leq N$ optima distincts de l'échantillon original de taille N), il est opportun de poursuivre l'exploration heuristique de l'espace des solutions possibles, ou au contraire s'il est plus judicieux de s'arrêter. En d'autres termes, le problème de la proximité du meilleur optimum local atteint vis-à-vis de l'optimum global est lié au problème de choix qui consiste à décider s'il est utile de poursuivre

l'exploration (essayer d'engendrer un meilleur optimum local), ou s'il faut s'arrêter.

6.1. UNE REGLE D'ARRET ELEMENTAIRE

Si $x^h \equiv X(\rho^h)$ est la meilleure solution parmi les N optima locaux, l'estimation $\hat{x}^*$ de l'optimum global fournit directement une approximation du potentiel d'amélioration $PA(\rho^h)$ par rapport à cette "meilleure" permutation ρ^h:

$$PA(\rho^h) = X(\rho^h) - x^* \simeq x^h - \hat{x}^*$$

Si C_o représente l'utilité, le bénéfice, le "coût d'opportunité" d'une unité gagnée sur la valeur du plan (c'est-à-dire le bénéfice que constitue la réduction d'une unité du coût $X(\rho)$ d'un plan ρ), la grandeur $PA(\rho^h).C_o$ représente l'utilité du gain maximum possible; si C_S représente le coût de calcul d'une exploration supplémentaire, la règle suivante est une première façon de décider de l'arrêt ou de la continuation de l'exploration:

si $(x^h - \hat{x}^*).C_o \geq C_S$ $\Rightarrow$ poursuivre l'exploration

si $(x^h - \hat{x}^*).C_o \leq C_S$ $\Rightarrow$ stop.

6.2. LA PROBABILITE D'AMELIORATION

La règle d'arrêt formulée ci-dessus est loin d'être satisfaisante: même si le potentiel d'amélioration est important, la probabilité d'amélioration, à savoir la probabilité qu'une exploration additionnelle aboutisse à une permutation meilleure que ρ^h, peut être très faible.

Ce qui nous intéresse, c'est la loi décrivant la probabilité d'engendrer, à partir d'une solution initiale aléatoire, un optimum local meilleur qu'une valeur donnée x^h. Ce qui nous intéresse, c'est la distribution définie par Reiter et Sherman, à savoir la distribution "induite" — voir paragraphe 1.3 et section 3 — décrivant le comportement du hill-climbing et de la procédure de génération aléatoire des solutions initiales. Dans la mesure où il est permis de considérer que la distribution Weibull "proprement dite" des optima locaux estimée à partir des R optima distincts[1] de l'échantillon original de taille N est simultanément une estimation de la distribution "induite" en question, c'est-à-dire une estimation de la loi de probabilités décrivant la distribution des optima susceptibles d'être engendrés, la probabilité d'amélioration qui nous intéresse est directement fournie par

$$\text{Prob}\ (X \leq x^h) = \hat{\Phi}(x^h)$$

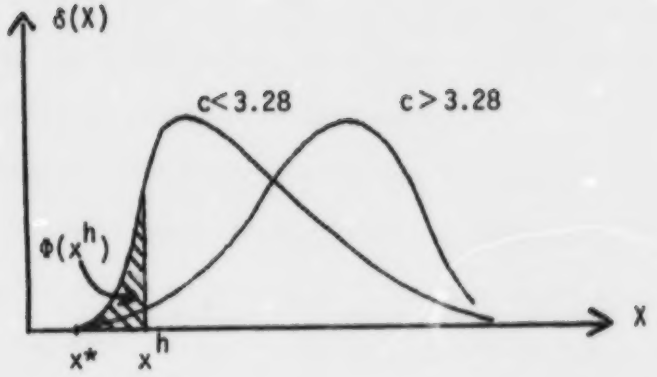
où $\hat{\Phi}(x^h)$ est la distribution Weibull estimée à partir des R optima distincts.

Rappelons que l'équivalence supposée ici entre la distribution "proprement dite" des optima locaux et la loi de probabilités décrivant la distribution "induite" n'est justifiée que pour les problèmes de grande taille, pour lesquels on a $R \simeq N$.

McRoberts [33] a mis en évidence le fait que la relation entre le potentiel d'amélioration et la probabilité d'amélioration dépend essentiellement du paramètre de forme c de la distribution.

---

1. Moyennant bien sûr la vérification préalable de l'hypothèse Weibull pour ces R optima distincts. Dans la suite de cette section, on continuera à supposer que cette hypothèse est vérifiée.



6.3. LA VALEUR ESPEREE DU GAIN ET LA REGLE DE McROBERTS

Un troisième critère doit jouer un rôle: il s'agit de la valeur espérée du gain fourni par une observation (une exploration) additionnelle. Les premiers à avoir introduit ce critère (ils furent d'ailleurs les premiers à discuter de l'élaboration d'une règle d'arrêt) sont Reiter et Sherman [35]. Leur papier est malheureusement assez obscur quant à la façon de calculer cette valeur espérée (ils n'avaient aucune idée de la distribution des optima locaux). McRoberts [33], par contre, a formulé une règle très simple en calculant la valeur espérée du gain de la façon suivante:

$$E(\text{Gain}) \approx \int_{\hat{x}^*}^{x^h} (x^h - X)\hat{\delta}(X)dX$$

où $\hat{\delta}(X)$ est la fonction de densité de la loi Weibull estimée à partir des $R \simeq N$ optima locaux distincts[1]. (Le gain est nul pour les valeurs de X supérieures ou égales à x^h, c'est ce qui explique la limite supérieure de l'intervalle d'intégration). La règle que propose McRoberts s'énonce comme suit:

---

1. Rappelons cependant que McRoberts n'avait pas fait la distinction — voir section 3 — entre la distribution "proprement dite" et la distribution "induite" des optima locaux.

si $E(Gain).C_0 < C_S$ => STOP

si $E(Gain).C_0 \geq C_S$ => poursuivre l'exploration

6.4. ELABORATION D'UNE REGLE GENERALE INTEGRANT LES TROIS CRITERES

L'inconvénient que présente la règle d'arrêt de McRoberts, c'est que même si la valeur espérée du gain sur une observation additionnelle est grande relativement à C_S/C_0,

a) l'obtention d'un optimum local effectivement meilleur que x^h peut nécessiter un nombre non négligeable d'observations (explorations) additionnelles; il serait donc judicieux que la valeur espérée du gain soit calculée non pas seulement vis-à-vis d'une observation additionnelle, mais plutôt vis-à-vis de plusieurs observations additionnelles, par exemple K, et que d'autre part l'évaluation des chances d'amélioration soit également réalisée sur ces K observations; cette dernière évaluation sera fournie par la probabilité qu'au moins une des K observations soit meilleure que x^h;

b) cette règle ne met pas en relation ladite valeur espérée du gain avec le potentiel d'amélioration $(\hat{x}^* - x^h)$; celui-ci intervient directement dans les limites d'intégration de l'intégrale de McRoberts, mais la règle d'arrêt de ce dernier n'établit aucune relation de comparaison entre les deux grandeurs en question; en fait, la valeur espérée du gain n'a vraiment de signification que relativement au gain maximum possible, à savoir le potentiel $(x^* - x^h)$ estimé par $(\hat{x}^* - x^h)$.

La suite de ce paragraphe propose une règle itérative

- définissant le nombre d'optima locaux qu'il faut engendrer à chaque itération pour contrôler simultanément la probabilité d'amélioration sur chacune de ces suites d'observations.
- définissant des conditions d'arrêt basées sur la comparaison de la valeur espérée du gain sur la suite d'observations de l'itération subséquente avec le potentiel d'amélioration et avec le coût total de calcul de cette itération subséquente.

1) Calcul de la probabilité d'amélioration sur K observations additionnelles

La probabilité d'amélioration dans une "épreuve" (une exploration) additionnelle étant fournie par $\hat{\Phi}(x^h)$, dans K épreuves la densité de probabilité de la variable aléatoire Z (Z = nombre de fois qu'une amélioration apparaît dans K épreuves, c'est-à-dire: nombre d'observations $\leq x^h$ dans K observations additionnelles) est fournie par la loi binomiale

$$f(Z) \approx \binom{K}{Z} [\hat{\Phi}(x^h)]^Z [1-\hat{\Phi}(x^h)]^{K-Z} \qquad Z=0,1,2,\ldots,K$$

Il résulte que la probabilité d'amélioration sur K observations est donnée par la probabilité d'avoir au moins une des K observations $\leq x^h$, à savoir

$$1 - \text{Prob}\ (Z=0) = 1 - [1-\hat{\Phi}(x^h)]^K$$

Si on fixe cette probabilité à une valeur α (par exemple α=0.80), on en déduit immédiatement le nombre d'observations additionnelles qui sont nécessaires pour garantir ce niveau de probabilité: ce nombre doit être au moins égal à la valeur de K obtenue en résolvant l'équation

$$1 - [1 - \hat{\Phi}(x^h)]^K = \alpha \qquad K=0,1,2,\ldots$$

ou

$$K = \frac{\log(1-\alpha)}{\log[1-\hat{\Phi}(x^h)]} \qquad K=0,1,2,\ldots$$

Cette procédure rend donc possible le choix du nombre d'observations additionnelles qui contrôle la probabilité d'amélioration.

2) Calcul de la valeur espérée du gain sur K observations additionnelles

Cette grandeur est fournie par la valeur espérée du gain maximum sur K observations additionnelles, c'est-à-dire par la valeur espérée de la va-

riable aléatoire $x^h - X_{(1)}$ où $X_{(1)}$ représente la plus petite valeur observée dans un échantillon de taille K issu de la population constituée des optima locaux, une population dont la distribution n'est rien d'autre que la distribution "proprement dite" dont il fut question précédemment, c'est-à-dire une population de densité $\delta(x)$ (Weibull). La densité de probabilité de la variable $X_{(1)}$ est donc la densité marginale de la plus petite variable statistique ordonnée correspondant à un échantillon de taille K issu de $\delta(x)$; cette fonction de densité est [7]:

$$p_{X_{(1)}}(X) \approx K\,[1-\hat{\Phi}(X)]^{K-1}\,\hat{\delta}(X)$$

La valeur espérée du gain sur K observations est donc estimée par:

$$E_K(\text{Gain}) \approx \int_{\hat{x}^*}^{x^h} (x^h - X)\,p_{X_{(1)}}(X)\,dX$$

$$\approx K \int_{\hat{x}^*}^{x^h} (x^h - X)[1-\hat{\Phi}(X)]^{K-1}\hat{\delta}(X)\,dX$$

L'estimateur de la valeur espérée du gain chez McRoberts est un cas particulier (K=1) de l'estimateur proposé ci-dessus.

3) Description de la procédure

La procédure itérative développée ci-dessous décrit la démarche complète pour engendrer successivement des ensembles de N_1, N_2, ... optima locaux.

Itération 1

- Générer N_1 optima locaux. Soit $x^{1,h}$ la valeur de la meilleure permutation. Soit R_1 le nombre d'optima distincts.
- Estimer les trois paramètres de la loi Weibull à laquelle on identifie la distribution [1] des R_1 optima locaux distincts $\Rightarrow \hat{a}_1, \hat{b}_1, \hat{c}_1$.

1. Rappelons encore une fois que, dans le cas des problèmes de grande taille où $R_1 \simeq N_1$, les estimations de la loi Weibull à partir des N_1 optima originaux et à partir des R_1 optima distincts conduisent quasiment aux mêmes résultats.

- Choisir le niveau de probabilité d'amélioration que l'on veut garantir sur les N_2 observations éventuelles de l'itération 2; soit α_1, ce niveau.
- Calculer N_2 en résolvant l'équation

$$N_2 = \frac{\log(1-\alpha_1)}{\log[1-\hat{\Phi}_1(x^{1,h})]}$$

où $\hat{\Phi}_1(X)$ est la loi Weibull avec les valeurs $\hat{a}_1$, $\hat{b}_1$ et $\hat{c}_1$ pour les paramètres.

- Calculer la valeur espérée du gain sur les N_2 observations additionnelles

$$E_{N_2}(\text{Gain}) = N_2 \int_{\hat{x}^*}^{x^{1,h}} (x^{1,h}-X)\,[1-\hat{\Phi}_1(X)]^{N_2-1}\,\hat{\delta}_1(X)dX \quad \text{où} \quad \hat{x}^* = \hat{a}_1$$

- Si l'une ou l'autre des deux conditions suivantes est remplie, arrêter l'exploration; sinon, passer à l'itération 2.

a) $E_{N_2}(\text{Gain}) < \beta_1 . PA(x^{1,h})$

$$< \beta_1(\hat{x}^* - x^{1,h})$$

où β_1 est la proportion limite (minimale) du potentiel d'amélioration définissant le seuil $\beta_1(\hat{x}^* - x^{1,h})$ que l'espérance du gain devrait dépasser pour suggérer la poursuite de l'exploration. (L'indice 1 associé au paramètre β signifie qu'une autre proportion limite pourrait être choisie à l'itération 2.)

b) $E_{N_2}(\text{Gain}).C_o < C_S.N_2 + C_E$

où C_o a été défini précédemment et représente le coût d'opportunité d'une unité de gain;

C_S a également été défini précédemment; il représente le coût d'une exploration additionnelle;

C_E est le coût correspondant à l'estimation des trois paramètres de la loi Weibull, estimation qui suivrait la génération des N_2 optima locaux supplémentaires.

Itération 2

- Générer N_2 optima locaux supplémentaires. Soit $x^{2,h}(\leq x^{1,h})$ la valeur de la meilleure permutation sur les deux itérations, c'est-à-dire sur les N_1+N_2 observations. Soit R_2 le nombre d'optima distincts parmi les N_2 nouveaux optima engendrés.
- Estimer la loi Weibull à partir des R_1+R_2 optima locaux distincts => $\hat{a}_2$, $\hat{b}_2$ et $\hat{c}_2$.
- Choisir le niveau de probabilité d'amélioration que l'on veut garantir sur les N_3 observations éventuelles de la troisième itération; soit α_2 ce niveau.
- Calculer N_3 en résolvant l'équation

$$N_3 = \frac{\log(1-\alpha_2)}{\log[1-\hat{\Phi}_2(x^{2,h})]}$$

où $\hat{\Phi}_2(X)$ est la loi Weibull avec les valeurs $\hat{a}_2$, $\hat{b}_2$ et $\hat{c}_2$ pour les paramètres.

- Calculer la valeur espérée du gain sur les N_3 observations éventuelles

$$E_{N_3}(\text{Gain}) = N_3 \int_{\hat{x}^*}^{x^{2,h}} (x^{2,h}-X)\,[1-\hat{\Phi}_2(X)]^{N_3-1}\,\hat{\delta}_2(X)dX \quad \text{où} \quad \hat{x}^* = \hat{a}_2$$

- Si $E_{N_3}(\text{Gain}) < \beta_2(\hat{x}^* - x^{2,h})$

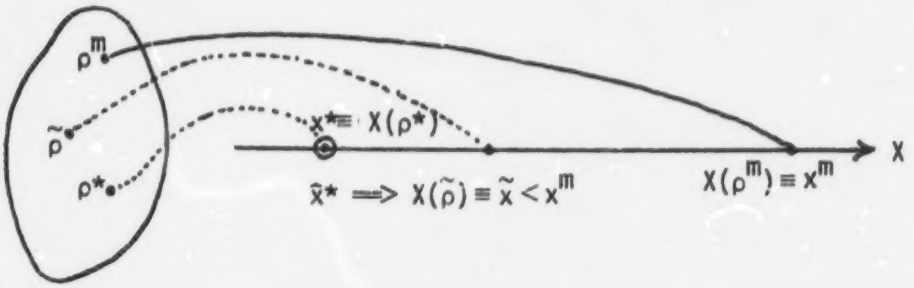
  ou si $E_{N_3}(\text{Gain}).C_o < C_S.N_3 + C_E$

  ] arrêter l'exploration

Sinon, passer à l'itération 3. Etc...

7. UN ALGORITHME HEURISTIQUE A PRIORI, PARFOIS EXACT A POSTERIORI

Les estimations simples de l'optimum global et les estimations d'intervalles de confiance constituent des informations qui pourraient être mises à profit pour explorer plus efficacement l'espace des solutions possibles. Ces estimations fournissent en effet une sorte "d'indication de direction" à suivre pour s'orienter vers l'optimum global ou vers les optima locaux qui lui sont proches. En d'autres termes, si $\hat{x}^*$ est une estimation simple de l'optimum global, l'objectif est d'utiliser $\hat{x}^*$ pour découvrir une permutation $\tilde{\rho}$ meilleure que la meilleure permutation connue, que l'on peut noter ρ^m.



7.1. DESCRIPTION DE LA PROCEDURE

Dans ce qui suit, le symbole $\rho^S(x)$ servira à désigner une permutation ayant le coût x.

Soit ρ^0 la meilleure permutation qui a été engagée dans la procédure d'estimation ayant abouti à $\hat{x}^*$.

- Si on utilise l'estimation $\hat{x}_1^*$ du niveau 1 (paragraphe 5.1), on a:

$$\rho^0 = \rho^S\,(x_{(1)})$$

où $x_{(1)}$ est la meilleure permutation parmi N permutations aléatoires.

- Si on utilise l'estimation $\hat{x}^*_{2,a}$, ou $\hat{x}^*_{2,b}$ ou encore $\hat{x}^*_{2,c}$ du niveau 2, on a:

$$\rho^0 = \rho^S(x^h)$$

où x^h est le meilleur parmi les optima locaux engendrés par un algorithme de hill-climbing, à partir des N solutions aléatoires.

- Si on utilise l'estimation $\hat{x}^*_3$ du niveau 3, on a:

$$\rho^0 = \rho^S(x^h)$$

- Si on utilise l'estimation $\hat{x}^*_4$ du niveau 4, on a:

$$\rho^0 = \rho^S(x^h)$$

où x^h est le meilleur parmi les optima locaux engendrés par différents algorithmes.

- Si on utilise l'estimation $\hat{x}^*_5$ du niveau 5, on a:

$$\rho^0 = \rho^S(x^{h,(1)}_{(1)})$$

où $x^{h,(1)}_{(1)}$ est le plus petit des E minima de E échantillons (de taille m) d'optima locaux engendrés par un algorithme de hill-climbing.

- Si on utilise l'estimation $\hat{x}^*_6$ du niveau 6, on a:

$$\rho^0 = \rho^S(x^{h,(1)}_{(1)})$$

où $x^{h,(1)}_{(1)}$ est le meilleur parmi les E m optima locaux produits par différents algorithmes.

Si des estimations ont été réalisées sur plusieurs niveaux, on possède donc une permutation ρ^0 pour chacun de ces niveaux. Soit ρ^m la meilleure permutation connue, c'est-à-dire la meilleure des permutations ρ^0. Bien

sûr, une procédure d'estimation ayant abouti à une estimation $\hat{x}^*$ telle que $\hat{x}^* \geq X(\rho^m)$ n'est d'aucune utilité.

L'objectif est d'utiliser l'une ou l'autre estimation simple $\hat{x}^* < X(\rho^m)$ pour dériver une permutation $\tilde{\rho}$ meilleure que la meilleure permutation connue, c'est-à-dire telle que:

$$X(\rho^*) \equiv x^* \leq X(\tilde{\rho}) \equiv \tilde{x} \leq X(\rho^m) \equiv x^m$$

En réalité, l'estimation $\hat{x}^*$ peut être supérieure, égale ou inférieure à l'optimum global, et on ne dispose d'aucune information à ce sujet.
L'idée est

- de supposer que l'estimation $\hat{x}^*$ est supérieure ou égale à x^*, et donc qu'elle constitue un meilleur majorant de x^* que ne l'est $X(\rho^m)$

- d'engager ce "majorant" $\hat{x}^*$ comme borne supérieure initiale dans un algorithme de branch and bound; cet algorithme devient alors une heuristique puisqu'on n'a pas la garantie que $\hat{x}^*$ est vraiment un majorant, et par conséquent, on ne peut garantir que la solution générée (s'il y en a une) est la solution exacte; nous allons cependant montrer que cette procédure, heuristique à priori, devient parfois (au vu de ses outputs) exacte à posteriori.

Quels sont les outputs possibles lorsqu'on utilise une estimation $\hat{x}^* < X(\rho^m)$ comme "majorant" initial dans un algorithme de branch and bound?

1) Ou bien la réalité est $x^* \leq \hat{x}^*$: $\hat{x}^*$ est effectivement un majorant. Dans ce cas, la procédure doit se dérouler d'une façon presque [1] normale.

1. La seule différence avec un branch and bound "classique" est que, ici, le majorant initial n'est pas la valeur de la meilleure permutation initialement connue.

Dès qu'on aboutit à l'extrémité d'une branche, c'est-à-dire en un sommet ne contenant plus qu'une permutation, et que la borne inférieure [1] associée à ce sommet a une valeur inférieure ou égale au majorant, on ajuste celui-ci et la "meilleure permutation connue jusque-là":

- cette "meilleure permutation connue jusque-là", qui initialement est ρ^m, est remplacée par la permutation "terminale" qu'on vient de trouver;
- le majorant, qui initialement est $\hat{x}^*$, est remplacé par la valeur de cette permutation "terminale".

La procédure aboutit nécessairement à l'optimum global: $X(\tilde{\rho}) \equiv x^*$; à noter que la permutation $\tilde{\rho}$ qu'on obtient est nécessairement telle que $X(\tilde{\rho}) \leq \hat{x}^*$.

2) Ou bien la réalité est $\hat{x}^* < x^*$: $\hat{x}^*$ est en fait un minorant. Deux cas sont possibles.

a) Toutes les bornes inférieures associées aux branches successives de l'arbre sont supérieures au "majorant" initial $\hat{x}^*$. Cette situation n'est effectivement possible que si $\hat{x}^*$ est un minorant. On est averti de cette situation par le fait que toutes les branches sont rejetées.

b) Certaines bornes inférieures associées à certaines branches s'avèrent inférieures ou égales à $\hat{x}^*$. La procédure se poursuit; si certaines bornes inférieures continuent à rester inférieures ou égales au "majorant" $\hat{x}^*$, il est possible qu'on arrive à l'extrémité d'une branche (en un sommet ne contenant plus qu'une permutation): un branch and bound "classique" entraînerait l'élimination de cette permutation "terminale" puisque la valeur de celle-ci (la borne inférieure [1]) est nécessairement supérieure au "majorant" $\hat{x}^*$; ici, on compare d'abord la valeur de cette

1. Sans perte de généralité, on suppose ici que la borne inférieure associée à l'extrémité d'une branche s'identifie avec la valeur de la permutation correspondant à cette extrémité.

permutation terminale avec la valeur de la "meilleure permutation connue jusque-là" (initialement ρ^m): si la permutation "terminale" qu'on vient de trouver est meilleure que la "meilleure permutation connue jusque-là", elle la remplace; le majorant reste $\hat{x}^*$ puisque celui-ci étant en réalité un minorant , la permutatation "terminale" en question ne peut avoir une valeur inférieure ou égale à $\hat{x}^*$, elle ne peut donc constituer un meilleur "majorant" que $\hat{x}^*$. Si au contraire la permutation "terminale" envisagée est moins bonne que la "meilleure permutation connue jusque-là", elle est éliminée.

En d'autres termes, la différence essentielle avec un branch and bound "classique" se produit lors des phases d'élimination et d'ajustement au niveau de l'extrémité des branches (là où il n'y a plus qu'une seule solution). Dans un branch and bound "classique", la phase d'élimination, à ce niveau, consiste à rejeter toute permutation "terminale" dont la borne inférieure associée est supérieure au majorant et la phase d'ajustement consiste, s'il y a lieu, à ajuster la "meilleure permutation connue jusque-là" ainsi que le majorant (qui est constamment égal à la valeur de la "meilleure permutation connue jusque-là"). Ici, lorsque la phase de séparation a mis à jour des extrémités de branche et que les bornes inférieures associées sont supérieures à $\hat{x}^*$, on peut être dans le cas 1 ou le cas 2 ci-dessus; il faut alors

- n'éliminer une permutation "terminale" que si elle est moins bonne que la "meilleure permutation connue jusque-là";
- n'ajuster (s'il y a lieu) que la "meilleure permutation connue jusque-là".

Cette façon de procéder se poursuit aussi longtemps qu'on n'a pas trouvé une permutation "terminale" dont la borne inférieure associée est inférieure ou égale

au majorant initial $\hat{x}^*$, auquel cas on est certain que ce $\hat{x}^*$ est en réalité un majorant (cas 1 ci-dessus); la procédure se poursuit alors de façon "classique".

La figure No. 2 résume ces considérations. (MP = "meilleure permutation connue jusque-là").

En conclusion, il y a trois possibilités pour les outputs:

- Aucune solution n'est engendrée (toutes les branches de l'arbre ont été rejetées). On peut assurer que $\hat{x}^*$ est un minorant: $\hat{x}^* < x^*$.
- Une permutation $\tilde{\rho}$ (la dernière "meilleure permutation connue jusque là") est engendrée
  - Si $X(\tilde{\rho}) \leq \hat{x}^*$, la procédure a été exacte et on peut assurer que $\tilde{\rho} \equiv \rho^*$.
  - Si $X(\tilde{\rho}) > \hat{x}^*$, on a une solution heuristique meilleure que la solution de départ ρ^m; d'autre part, on peut assurer que $\hat{x}^*$ est un minorant: $\hat{x}^* < x^*$.

Bien sûr, il se peut aussi que la place de mémoire requise ou le temps d'exécution imposent l'arrêt prématuré de l'exploration.

En résumé, en utilisant l'estimation $\hat{x}^*$ comme "majorant" initial dans un algorithme de branch and bound (quelque peu adapté aux considérations qui précèdent), on s'engage dans un nouvel algorithme, heuristique a priori, mais parfois exact a posteriori.
Remarquons que, au lieu d'utiliser une estimation simple $\hat{x}^*$, il serait possible d'utiliser la borne inférieure d'un intervalle de confiance: on choisirait donc ce que l'on croit être une borne inférieure comme "majorant" initial dans le branch and bound; ce "majorant" étant fort probablement un minorant

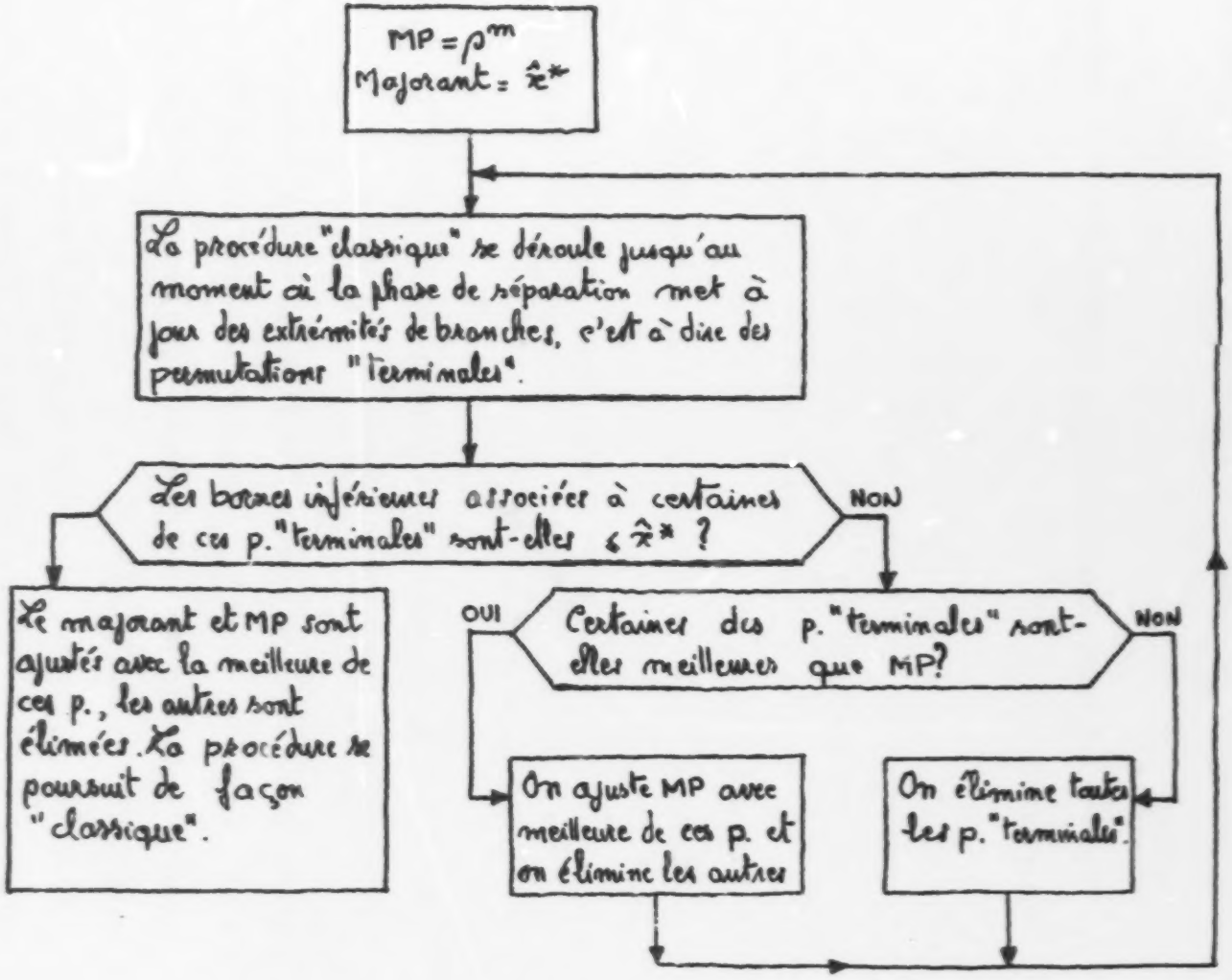

MP = ρ^m
Majorant = $\hat{x}^*$
La procédure "classique" se déroule jusqu'au moment où la phase de séparation met à jour des extrémités de branches, c'est à dire des permutations "terminales".
Les bornes inférieures associées à certaines de ces p. "terminales" sont-elles < $\hat{x}^*$?
NON
Le majorant et MP sont ajustés avec la meilleure de ces p., les autres sont élimées. La procédure se poursuit de façon "classique".
OUI
Certaines des p. "terminales" sont-elles meilleures que MP?
NON
On ajuste MP avec meilleure de ces p. et on élimine les autres
On élimine toutes les p. "terminales".

FIGURE No. 2

la procédure devrait alors rester heuristique a posteriori.

7.2. CALCUL D'INTERVALLES DE CONFIANCE

Lorsque la procédure décrite ci-dessus ne produit aucune solution (c'est-à-dire lorsque toutes les branches de l'arbre sont rejetées) ou bien lorsqu'elle engendre une permutation $\tilde{\rho}$ telle que $X(\tilde{\rho}) > \hat{x}^*$, on peut conclure que l'estimation simple $\hat{x}^*$ est en réalité une estimation par défaut, c'est-à-dire un minorant: $\hat{x}^* < x^*$. Cette conclusion constitue, en soi, une nouvelle information non négligeable puisqu'on en déduit alors très simplement un nouvel intervalle de confiance à 100% pour l'optimum global:

- Si on a engendré $\tilde{\rho}$ tel que $X(\tilde{\rho}) > \hat{x}^*$, on a

  $$\text{Prob}\ \{\hat{x}^* < x^* \leq X(\tilde{\rho})\} = 1$$

- Si on n'a engendré aucune solution, on a

  $$\text{Prob}\ \{\hat{x}^* < x^* \leq X(\rho^m)\} = 1$$

On obtient des intervalles de confiance du même type si, au lieu d'une estimation simple $\hat{x}^*$, on utilise la borne inférieure d'un intervalle de confiance comme "majorant" initial dans le branch and bound.

La figure N° 3 résume la démarche lorsqu'on engage une quantité θ comme "majorant" initial dans le branch and bound, que θ soit l'une ou l'autre estimation simple, ou encore la borne inférieure de l'un ou l'autre intervalle de confiance.

Corollaire

En fait, n'importe quelle quantité θ peut être supposée constituer un majorant de l'optimum global et être engagée comme "borne supérieure" initiale dans le branch and bound. Si on engendre une permutation $\tilde{\rho}$ telle

que $X(\tilde{\rho}) \leq \theta$, le procédure a fonctionné de façon exacte; si on n'engendre aucune solution, on déduit l'intervalle de confiance à 100%: $[\theta, X(\rho^m)]$; si on engendre $\tilde{\rho}$ tel que $X(\tilde{\rho}) > \theta$, on déduit l'intervalle de confiance à 100%: $[\theta, X(\tilde{\rho})]$. Cette procédure peut être utilisée plusieurs fois pour essayer d'engendrer, successivement, des intervalles de confiance à 100% de plus en plus étroits: à chaque itération i, on s'engage dans le branch and bound avec un "majorant" θ_i quelque peu supérieur à celui de l'itération précédente, et on ajuste la "meilleure permutation connue jusque-là" (notée MP dans ce qui suit); ainsi donc, on pourrait par exemple obtenir la suite d'itérations décrite ci-dessous.

Itération 1

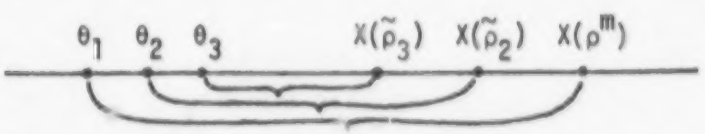
Majorant = θ_1
MP = ρ^m $\longrightarrow$ aucune solution engendrée $\longrightarrow$ $[\theta_1, X(\rho^m)]$ à 100%

Itération 2

Majorant = $\theta_2 > \theta_1$
MP = ρ^m $\longrightarrow$ on engendre $\tilde{\rho}_2$ tel que $X(\tilde{\rho}_2) > \theta_2$ $\longrightarrow$ $[\theta_2, X(\tilde{\rho}_2)]$ à 100%

Itération 3

Majorant = $\theta_3 > \theta_2$
MP = $\tilde{\rho}_2$ $\longrightarrow$ on engendre $\tilde{\rho}_3$ tel que $X(\tilde{\rho}_3) > \theta_3$ $\longrightarrow$ $[\theta_3, X(\tilde{\rho}_3)]$ à 100%



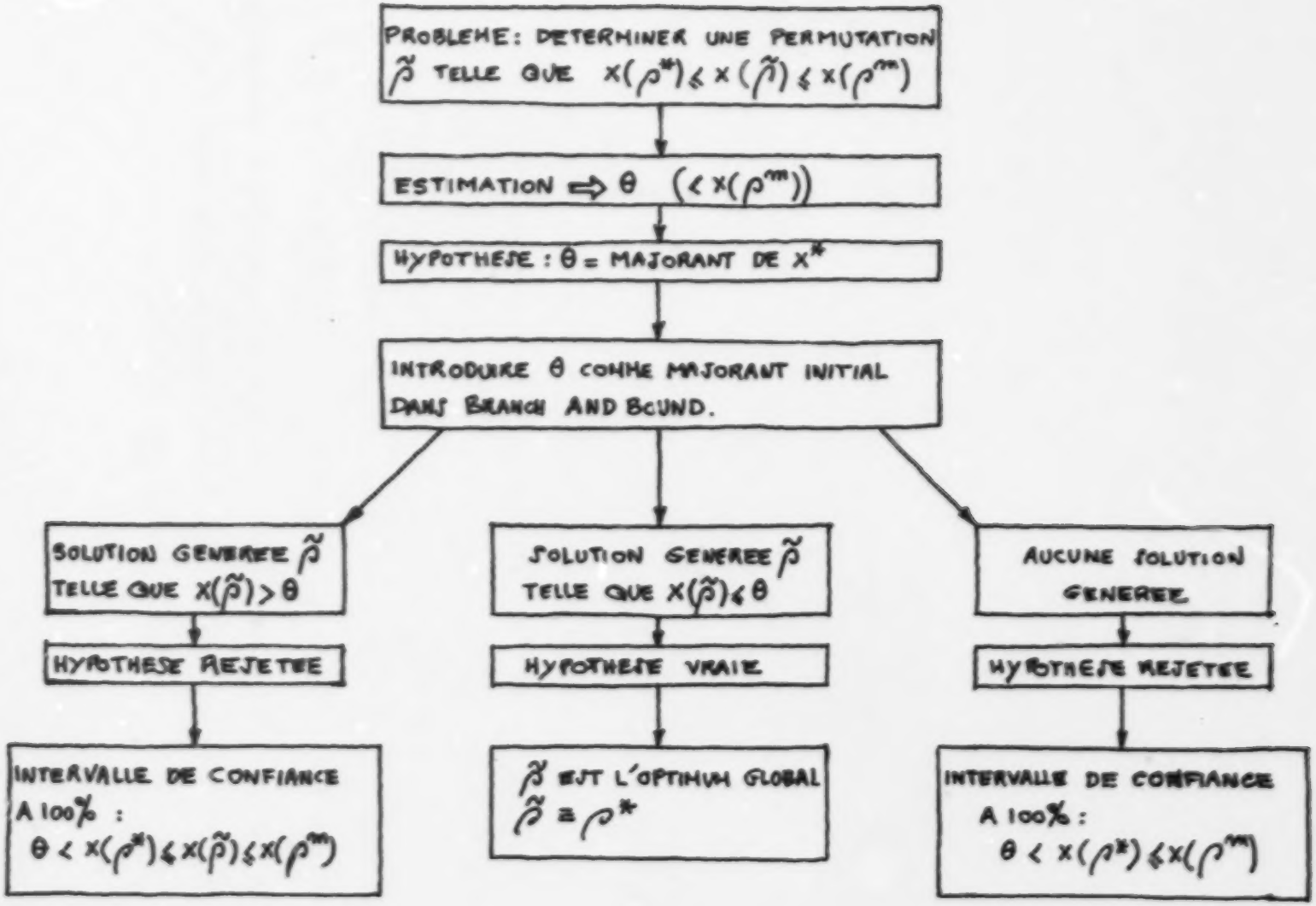

PROBLEME: DETERMINER UNE PERMUTATION $\tilde{\rho}$ TELLE QUE $X(\rho^*) \leq X(\tilde{\rho}) \leq X(\rho^{mm})$
ESTIMATION $\Rightarrow \theta$ $(< X(\rho^{mm}))$
HYPOTHESE : θ = MAJORANT DE X^*
INTRODUIRE θ COMME MAJORANT INITIAL DANS BRANCH AND BOUND.
SOLUTION GENEREE $\tilde{\rho}$ TELLE QUE $X(\tilde{\rho}) > \theta$
SOLUTION GENEREE $\tilde{\rho}$ TELLE QUE $X(\tilde{\rho}) < \theta$
AUCUNE SOLUTION GENEREE
HYPOTHESE REJETEE
HYPOTHESE VRAIE
HYPOTHESE REJETEE
INTERVALLE DE CONFIANCE A 100% : $\theta < X(\rho^*) \leq X(\tilde{\rho}) \leq X(\rho^{mm})$
$\tilde{\rho}$ EST L'OPTIMUM GLOBAL $\tilde{\rho} = \rho^*$
INTERVALLE DE CONFIANCE A 100% : $\theta < X(\rho^*) \leq X(\rho^{mm})$

FIGURE N° 3

CONCLUSION

L'approche développée dans ce rapport

1. pour estimer l'optimum global des problèmes combinatoires de grande taille (estimation nécessaire à l'évaluation de la qualité de toute solution heuristique).
   - soit à partir de solutions aléatoires, éventuellement combinées avec l'une ou l'autre solution heuristique
   - soit à partir de l'information contenue dans les optima locaux engendrés par une procédure de hill-climbing
2. pour choisir la taille des suites d'optima locaux qu'une procédure de hill-climbing doit engendrer successivement et pour fixer les conditions d'arrêt de cette exploration itérative
3. pour tirer parti, dans un branch and bound, de l'information résultant de l'estimation de l'optimum global

est suffisamment générale pour être appliquée à de nombreux problèmes d'optimisation combinatoire.

Pour ce qui est des deux premiers objectifs que s'était fixés ce rapport, la théorie de l'inférence statistique et la théorie des valeurs extrêmes constituent un outil extrêmement pertinent. Certaines hypothèses sont difficiles à justifier théoriquement de façon rigoureuse, mais la preuve qu'elles ne peuvent être conservées demeure cependant du ressort de l'expérience; jusqu'à preuve du contraire, elles doivent donner satisfaction.

Ce rapport illustre combien le recoupement de la statistique théorique et de l'optimisation mathématique devient important. Au niveau de la programmation mathématique, l'implication du travail qui a été présenté est

important en ce sens que les techniques heuristiques s'avèrent actuellement indispensables pour résoudre certains problèmes de grande taille, notamment ceux qui appartiennent à la famille "NP-complete".

ANNEXES

ANNEXE 1

TEST DE KOLMOGOROV-SMIRNOV [6,7]

Le test d'ajustement de K-S réalise la comparaison d'une fonction de répartition théorique $\Phi(Y)$ avec la fonction de répartition d'un échantillon (sample or empirical cumulative distribution function).

Plaçons-nous dans un cas général où le test porte sur l'ajustement de T observations y_i, non nécessairement toutes distinctes. (C'est le cas des niveaux 5 et 6 du paragraphe 5.1; c'est également le cas lorsqu'on veut tester l'ajustement de N optima locaux initialement engendrés à la distribution Weibull estimée à partir des R optima distincts, test auquel il fut fait allusion dans la section 3 pour vérifier si l'estimation de la distribution "proprement dite" des optima peut également être considérée comme une estimation de la loi de probabilités décrivant le comportement du hill-climbing et de la procédure de génération aléatoire des solutions initiales).

La fonction de répartition empirique de l'échantillon des T observations $y_1, y_2, \ldots, y_T$ est simplement

$$S_T(Y) = \frac{\text{nombre d'observations } y_i \ (i=1,\ldots,T) \leq Y}{T}$$

Soient $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(T)}$ les observations $y_1, y_2, \ldots, y_T$ ordonnées par ordre de valeurs croissantes.
Soient $y^d_{(1)} < y^d_{(2)} < \ldots < y^d_{(R)}$ cette suite d'observations ordonnées après suppression des observations identiques, c'est-à-dire la suite des observations ordonnées distinctes ($R \leq T$).

La fonction de répartition représentant l'échantillon $y_1, y_2, \ldots, y_T$ est

$$S_T(Y) = \begin{cases} 0 & \text{si } Y < y^d_{(1)} \\ (\sum_{i=1}^{k} f_i)/T & \text{si } y^d_{(k)} \leq Y < y^d_{(k+1)} \text{ avec } k=1,2,\ldots,R-1 \\ 1 & \text{si } Y \geq y^d_{(R)} \end{cases}$$

où f_i représente le nombre de fois qu'apparaît la valeur $y^d_{(i)}$ dans l'échantillon de départ $y_1, y_2, \ldots, y_T$.

La statistique (le critère) utilisée dans ce test est

$$D_T = \underset{Y \in P}{SUP} \; | \; S_T(Y) - \Phi(Y) \; |$$

où $\begin{cases} P = \{y^d_{(1)}, y^d_{(2)}, \ldots, y^d_{(R)}\} \\ SUP = \text{plus petite borne supérieure} \end{cases}$

$$\Rightarrow D_T = \underset{k=1,2,\ldots,R}{Max} \; [\, |S_T(y^d_{(k-1)}) - \Phi(y_{(k)})|, |S_T(y^d_{(k)}) - \Phi(y_{(k)})| \,]$$

$$= \underset{k=1,2,\ldots,R}{Max} \; [\, |(\sum_{i=1}^{k-1} f_i)/T - \Phi(y_{(k)})|, |(\sum_{i=1}^{k} f_i)/T - \Phi(y_{(k)})| \,]$$

où pour des raisons de cohérence dans les notations, on doit poser

$S_T(y^d_{(0)}) \equiv (\sum_{i=1}^{0} f_i)/T \equiv 0.$

La distribution d'échantillonnage de la variable D_T étant connue et tabulée, on peut déterminer la valeur critique $D_{T,\alpha}$ telle que

$$\text{Prob } (D_T \geq D_{T,\alpha}) = \alpha \qquad \text{(probabilité d'erreur de type I)}$$

Le test de K-S, au niveau de signification α, consiste, lorsque le D_T observé est $\geq D_{T,\alpha}$ à rejeter l'hypothèse nulle selon laquelle les observations $y_1, y_2, \ldots, y_T$ s'ajustent à la distribution théorique $\Phi(Y)$.

Remarquons que si le test porte sur l'ajustement de T observations toutes distinctes les unes des autres (R ≡ T), on a plus simplement encore [1]

$$D_T = \max_{k=1,2,\ldots,T} \left[\left| \frac{k-1}{T} - \Phi(y_{(k)}) \right|, \left| \frac{k}{T} - \Phi(y_{(k)}) \right| \right]$$

Etant donné que, dans notre cas

- la distribution théorique $\Phi(Y)$ n'est connue [2] qu'après estimation de ses paramètres à partir des mêmes observations qui forment la distribution empirique
- la distribution est discrète

le niveau de signification est en fait au plus égal à α et le D_T calculé est une estimation $\hat{D}_T$ dont la distribution d'échantillonnage est différente de celle de D_T; le test est approximatif.

---

1. C'est le cas des tests à réaliser aux niveaux 3 et 4 du paragraphe 5.1.

2. Sauf dans le cas du test d'ajustement de N optima locaux initiaux à la distribution Weibull des R optima distincts — Voir section 3 —. Dans ce cas, la distribution théorique doit être considérée comme une distribution "a priori".

ANNEXE 2

TESTS DE CARACTERE ALEATOIRE [7]

Dans le cas où, par exemple, l'hypothèse Weibull porte sur la distribution de R optima locaux distincts (paragraphe 5.1, niveaux 3 et 4), il s'agit de tester le caractère aléatoire de la suite de nombres x_1^{hd}, x_2^{hd}, ..., x_R^{hd}.

On classe les nombres en deux catégories, A et B. Par exemple

$$x_i^{hd} \in A \qquad \text{si } x_i^{hd} < x_M^R$$

$$x_i^{hd} \in B \qquad \text{si } x_i^{hd} > x_M^R$$

où x_M^R est la médiane des R observations. (La coutume semble être de ne pas prendre en considération les x_i^{hd} qui seraient égaux à la médiane). La suite de nombres devient donc une suite de R symboles AABABBBA.

1) Test basé sur le nombre total de "runs"

"A run is defined to be a succession of one or more identical symbols which are followed and preceded by a different symbol or no symbol at all" [7].

Soient T^A et T^B les nombres de "runs" de type A et B. Le nombre total de "runs" est $T = T^A + T^B$.

Dans l'hypothèse où la suite en question est aléatoire, la distribution d'échantillonnage de la variable T est connue et tabulée. Par conséquent, on peut déterminer les nombres critiques T_α^1 et T_α^2 correspondant à un niveau de signification au plus égal à α donné, c'est-à-dire tels que

$$\text{Prob } (T \leq T_\alpha^1 \text{ ou } T \geq T_\alpha^2) \leq \alpha \qquad \text{(probabilité d'erreur de type I)}$$

("au plus égal à α" car la variable T étant discrète, il n'y a qu'un

nombre fini de valeurs possibles pour la probabilité d'erreur de type I).

Le test basé sur le nombre total de "runs", au niveau de signification α, consiste à rejeter l'hypothèse du caractère aléatoire de la suite de nombres $h_1^{hd}, x_2^{hd}, \ldots, x_R^{hd}$ lorsque $T \leq T_\alpha^1$ ou $T \geq T_\alpha^2$

2) Test basé sur la longueur du plus long "run"

Dans l'hypothèse où la suite AABABBBA... correspondant à la suite $x_1^{hd}, x_2^{hd}, \ldots x_R^{hd}$ est aléatoire, la distribution d'échantillonnage de la variable V représentant la longueur du plus long "run" est connue et tabulée. Par conséquent, on peut déterminer la valeur critique V_α correspondant à un niveau de signification au plus égal à α donné, c'est-à-dire telle que

$$\text{Prob}\ (V \geq V_\alpha) \leq \alpha \qquad \text{(probabilité d'erreur de type I)}$$

Le test basé sur la longueur du plus long "run", pour le niveau de signification α, consiste à rejeter l'hypothèse du caractère aléatoire de la suite en question lorsqu'on obtient $V \geq V_\alpha$.

BIBLIOGRAPHIE

1. Clough, D. (1969) An Asymptotic Extreme Value Sampling theory for the Estimation of a Global Maximum. C O R S Journal, 7, 102-115.

2. Dannenbring, D. (1977) Estimating Optimal Solutions for Large Combinatorial Problems. Management Science, 23, 1273-1283.

3. Dionne, R. (1974) Une analyse théorique et numérique du problème de choix optimal du réseau de transport sans congestion. Publication #27, Centre de recherche sur les transports, Université de Montréal.

4. Fisher, M.L. (1978) Worst-Case Analysis of Heuristics; in Interfaces between Computer Science and Operations Research, Lenstra,J.K., Rinnooy Kan, A.H.G. and Van Emde Boas, P., Editors, Mathematical Centrum, Amsterdam.

5. Fisher, R. and Tippett, L. (1928) Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. Proc. Camb. Phil. Soc. 24, 180-191

6. Fisz, M. (1963) Probability Theory and Mathematical Statistics. John Wiley and Sons, Inc.

7. Gibbons, J.D. (1971) Non Parametric Statistical Inference. McGraw-Hill.

8. Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Mathematics, Vol. 44, No. 3, 423-453.

9. Golden, B. (1977) A Statistical Approach to the TSP. Networks, 7, 209-225.

10. Golden, B. and Alt, B. (1978) Interval Estimation of a Global Optimum for Large Combinatorial Problems. College of Business and Management, University of Maryland at College Park.

11. Gumbel, E. (1958) Statistics of Extremes. Columbia University Press.

12. Guldmann, J.M. (1973) Optimization Models of Air Pollution Control Strategies and Location of Industries. Research thesis for the degree of Master of Science in Urban and Regional Sciences, Israël Institute of Technology, Haïfa, Israël.

13. Guldmann,J.M. and Schefer, D. (1973) Mathematical Models of Industrial Plant Location and Abatement Strategies. Environment and Planning, Vol. 5, No.5, 577-588.

14. Harter, H. and Moore, A. (1965) Maximum Likelihood Estimation of the Parameters of the Gamma and Weibull Populations from Complete and from Censored Samples. Technometrics, 7, 639-643.

15. Hopkins, L.D. (1975) Optimum-Seeking Models for Design of Suburban Land Use Plans. Ph.D. Dissertation in City and Regional Planning, University of Pennsylvania.

16. Koopmans, T. and Beckmann, M. (1957) Assignment Problems and the Location of Economic Activities. Econometrica 25, no. 1, 53-76.

17. Lardinois, C. (1976) Pollution de l'environnement et planification du développement urbain et régional. Publication # 52, Centre de recherche sur les transports, Université de Montréal.

18. Lardinois, C. (1977) Les modèles de design en aménagement urbain et régional. Publication # 64, Centre de recherche sur les transports, Université de Montréal.

19. Lardinois, C. (1978) Pollution de l'air et modèles d'organisation spatiale. Publication # 73, Centre de recherche sur les transports, Université de Montréal.

20. Lenstra, J.K. and Rinnoy Kan, A.H.G. (1978) Computational Complexity of Discrete Optimization Problems; in Interfaces between Computer Science and Operations Research, Lenstra, J.K., Rinnoy Kan, A.H.G. and Van Emde Boas, P., Editors, Amsterdam, Mathematical Centrum.

21. Los, M. (1975) Simultaneous Optimization of Land Use and Transportation in New Town Design. Ph.D. Dissertation in Urban and Regional Planning, University of Pennsylvania.

22. Los, M. (1976) An Optimal Network Problem Without Congestion: Some Computational Results. Publication #40, Centre de recherche sur les transports, Université de Montréal.

23. Los, M. (1976) A Land Use Plan Design Model with Fixed Interactions and Central Facility Interactions. Publication #53, Centre de recherche sur les transports, Université de Montréal.

24. Los, M. (1977) Solution of the Network Design Problem with Congestion and Fixed Demands for Travel by an All-or-Nothing Traffic Assignment Algorithm. Publication #56, Centre de recherche sur les transports, Université de Montréal.

25. Los, M. (1977) Comparison of Several Heuristic Algorithms to Solve Quadratic Assignment Problems of the Koopmans-Beckmann Type. Publication #87, Centre de recherche sur les transports, Université de Montréal.

26. Los, M. (1978) Simultaneous Optimization of Land Use and Transportation. Regional Science and Urban Economics, 8, 21-42.

27. Los, M. (1979) A Discrete-Convex Programming Approach to the Simultaneous Optimization of Land-Use and Transportation. A paraître dans Transportation Research.

28. Los, M. and Hopkins, L.D. (1979) Location-Allocation Algorithms for Land Use Plans Design with Fixed and Substitutable Interactions. A paraître dans The Journal of Regional Science.

29. Los, M. and Nguyen, S. (1977) Spatial Allocation on a Network with Congestion. Publication #59, Centre de recherche sur les transports, Université de Montréal.

30. Lundqvist, L. (1973) Integrated Location-Transportation Analysis: A Decomposition Approach. Regional and Urban Economics, Vol. 3, No.3,233-262.

31. Lundqvist, L. (1977) Planning for Freedom of Action. Papier présenté à la International Research Conference on Spatial Interaction Theory and Planning Models, Bastad (Sweden).

32. Mann, N., Schafer, R. and Singpurwalla, N. (1974) Methods for Statistical Analysis of Reliability and Life Data. John Wiley and Sons.

33. McRoberts, K. (1971) A Search Model for Evaluating Combinatorially Explosive Problems. Operations Research, 19, 1331-1349.

34. Mood, A.M. and Graybill, F.A. (1963) Introduction to the Theory of Statistics. McGraw-Hill.

35. Reiter, S. and Sherman, G. (1965) Discrete Optimizing. J. Soc. Indust. Appl. Math., Vol. 13, No. 3, 864-889.

36. Revelle, C., Marks, D. and Liebman, J.C. (1970) An Analysis of Private and Public Sector Location Models. Management Science, Vol. 16, No. 11, 692-707.

37. Robson, D. and Whitlock, J. (1964) Estimation of a Truncation Point. Biometrika, 51, 33-39.

38. Roy, B. (1969-1970) Algèbre moderne et théorie des graphes, Tomes I et II, Dunod.

39. Sahni, S. (1977) General Technics for Combinatorial Approximation. Operations Research, 25. 920-936

40. Scott, A.J. (1969) Combinatorial Programming and the Planning of Urban and Regional Systems. Environment and Planning, 1, 125-142.

41. Scott, A.J. (1971) Dynamic Location-Allocation Systems: Some Basic Planning Strategies. Environment and Planning, 3, 73-82.

42. Weibull, W. (1951) A Statistical Distribution Function of Wide Applicability. J. Appl. Mech., 18, 293-297.

43. Weide, B. (1977) A Survey of Analysis Techniques for Discrete Algorithms. Computing Surveys, Vol. 9, No. 4, 291-313.

44. Zanakis, S. (1977) Computational Experience with Some Nonlinear Optimization Algorithms in Deriving Maximum Likelihood Estimates for the Three-Parameter Weibull Distribution. TIMS Studies in the Management Sc., 7, 63-77.